



Institut des
Politiques Publiques

RAPPORT IPP N° 35 - Mai 2022

Évaluation micro-économétrique du plan France Très Haut Débit

Paul Dutronc-Postel
Nicolas Ghio
Arthur Guillouzouic
Clément Malgouyres

A large, faint watermark of the IPP logo is positioned in the bottom right corner of the page. It consists of the lowercase letters 'ipp' in a bold, sans-serif font, centered within a teal-colored circular frame that is open at the top and bottom, matching the design of the main logo.



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

www.ipp.eu





RAPPORT IPP - Mai 2022

Évaluation micro-économétrique du plan France Très Haut Débit

Paul Dutronc-Postel
Nicolas Ghio
Arthur Guillouzouic
Clément Malgouyres

L'ÉQUIPE

Paul Dutronc-Postel rejoint l'IPP en septembre 2019 après une thèse en économie du développement au sein de l'École d'économie de Paris. Il y travaille désormais sur les problématiques liées à la fiscalité des ménages, au marché du travail, et aux entreprises. Page personnelle : <http://www.ipp.eu/annuaire/paul-dutronc/>

Nicolas Ghio diplômé du Master in Economics de Sciences Po, et après un passage à UC Berkeley, Nicolas Ghio rejoint l'IPP en septembre 2020. Il y travaille principalement sur des questions liées à la fiscalité des entreprises et au marché du travail. Page personnelle : <https://sites.google.com/view/nicolas-ghio/home>

Arthur Guillouzouic est économiste à l'Institut des politiques publiques (IPP). Ancien élève de l'École normale supérieure Paris-Saclay et titulaire d'un doctorat en sciences économiques de Sciences Po, ses recherches se situent dans les champs de l'économie publique et de l'économie de l'innovation.

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

Clément Malgouyres est économiste à l'Institut des politiques publiques (IPP) et chercheur associé à l'École d'économie de Paris (PSE). Il s'intéresse à des questions empiriques en économie publique, économie internationale et économie du travail. Ses travaux récents portent sur la fiscalité des entreprises, l'évaluation quantitative des politiques publiques et le déploiement des technologies ADSL et Très Haut Débit. Page personnelle : <https://sites.google.com/site/clementmalgouyres>

INTRODUCTION ET SYNTHÈSE DES RÉSULTATS

L'accès à l'internet à haut débit s'est développé rapidement dans le monde entier depuis la fin des années 1990. La diffusion de cette technologie a permis d'accélérer la production et la distribution d'information et d'idées et a favorisé la croissance économique en facilitant le développement et l'adoption d'innovations (Czernich et al., 2011). Si les effets positifs en matière économique de la première génération de l'internet à haut débit sont désormais bien documentés (voir par exemple Bertschek et al., 2015, pour une revue exhaustive), il existe moins de travaux empiriques sur les effets des réseaux plus récents permettant un accès au très haut débit (THD). L'investissement public et privé dans la diffusion de l'accès du THD est considérable. Le Plan France Très Haut Débit (PFTHD) s'inscrit dans un agenda européen dont la visée initiale est une couverture de 100 % des entreprises et des foyers européens par du très haut débit (supérieur à 30Mbt/s) d'ici 2022 et des débits supérieurs à 100 Mbt/s en 2025 (objectifs de la Gigabit Society). Le montant total de l'investissement en France avoisine les 20 milliards d'euros.

Un des objectifs affichés de cet effort considérable est de “fortifier le tissu des entreprises françaises [...], de favoriser l'émergence de leaders mondiaux, mais aussi de faire en sorte que tous les acteurs de l'économie s'approprient le numérique [...]”. La politique d'investissement dans le THD repose donc sur l'hypothèse d'un effet positif du THD sur la performance des entreprises et l'attractivité des

territoires. L'évaluation des effets socio-économiques des réseaux THD est donc cruciale afin de juger de l'opportunité de l'effort d'investissement consenti et d'estimer la capacité du PFTHD à atteindre ses objectifs en matière de promotion de la compétitivité de l'économie française et de l'usage du numérique.

Objectifs de l'étude. L'objectif de ce projet de recherche est de mener une analyse quantitative s'intéressant aux effets multidimensionnels induits par l'arrivée du THD dans les territoires, à la fois sur le comportement des entreprises et des ménages. Notre étude se concentre en particulier sur les modifications dans l'usage du numérique induite par le THD (en particulier au sein des entreprises), et vise dans un second temps à estimer les impacts de ces changements d'usage sur la structure de l'activité (créations d'entreprises, relocalisation d'établissements, dynamiques d'emploi des entreprises), et sur les revenus et dépenses des ménages (population, revenus non salariaux, loyers et prix de l'immobilier). Vu le contexte de crise sanitaire traversé pendant la durée de ce projet de recherche, une partie de ce rapport utilise des données très récentes pour estimer la résilience des zones connectées au THD dans un contexte de distanciation des interactions sociales et professionnelles.

Cette analyse s'appuie sur des outils micro-économétriques qui visent à estimer l'effet causal de l'installation du THD dans une zone définie à un niveau fin (alternativement la commune et la parcelle cadastrale), en la comparant avec d'autres zones similaires ne bénéficiant pas du THD à cette période. En particulier, la comparaison de ces deux types de zones avant et après l'arrivée du THD, nous permet de quantifier l'effet de cet événement.

Contexte institutionnel. Le PFTHD distingue deux grands types de zones géographiques. Au sein des zones très denses (ZTD), les fournisseurs d'accès nationaux assurent le déploiement de la fibre jusqu'à l'abonné -- avec potentiellement des accords de co-investissement. Dans les régions de plus faible densité, où il n'est pas immédiatement rentable pour les opérateurs de déployer la fibre jusqu'à

l'abonné, les collectivités interviennent directement via la création de réseaux d'initiative publique (RIP), qu'elles gèrent elles-mêmes ou dont elles confient la gestion à des entreprises spécialisées. L'évaluation économétrique des effets du THD sur les comportements et variables que nous décrivons ci-dessous est généralement menée en premier lieu sans restrictions géographiques, puis avec une focalisation sur les zones RIP uniquement. La mesure d'impact sur ces zones en particulier est essentielle, puisque ces territoires sont les principaux bénéficiaires de la dépense publique consentie dans le cadre du PFTHD.

Synthèse des résultats. Ce rapport contient trois chapitres. Le premier chapitre présente les avancées effectuées concernant l'évaluation du plan France Très Haut Débit au niveau communal. Le second chapitre se focalise sur la crise sanitaire liée au Covid-19 et vise à estimer dans quelle mesure la disponibilité locale du très haut débit (THD) a permis aux économies locales de mieux résister à la crise. Le troisième chapitre propose une analyse à un niveau spatial très fin (la parcelle cadastrale), et évalue les impacts de la connection au THD sur la population locale, les revenus des habitants, et le marché immobilier.

Chapitre 1 : Analyse des effets du déploiement de la fibre au niveau communal

- Ce chapitre s'intéresse aux effets du **déploiement de la fibre** sur la dynamique de la démographie des entreprises et du marché de l'immobilier.
- Le choix de se focaliser sur la fibre plutôt que sur une classe de débit (THD) est principalement motivé par une question de disponibilité de données en 2019 et 2020 qui soient comparables avec les données la période précédente (2015-2018).
- Après avoir mobilisé les différentes sources disponibles mesurant l'évolution du THD, il en ressort qu'au vu de la qualité de ces données une approche

se concentrant seulement sur le très haut débit fibre est préférable. En effet les données de l'Agence du numérique ne couvrant que la période 2013-2018T1, seules les données ma-connexion permettent d'appréhender l'évolution du très haut débit, toutes technologies confondues post-2017. Or, les taux de couverture THD par commune présents dans ces données semblent en grand partie incohérents avec les précédents taux observés en 2017 dans les données de l'Agence du numérique. Une approche utilisant uniquement les données Agence du numérique et ARCEP sur le déploiement de la fibre couvrant la période 2015-2020 a donc été privilégiée.

- Nous adoptons une approche par différence-de-différences au niveau commune ainsi qu'au niveau commune–secteur. Nous utilisons une **estimation statique** où la variable de traitement est **la part des locaux ayant accès à la fibre**. De plus, nous estimons **une étude d'évènement dynamique** autour de l'année de forte hausse de ce taux de couverture.
- Il est important de préciser que les associations statistiques entre fibre et les variables d'intérêt considérées n'ont pas nécessairement de valeur causale dans la mesure où les décisions de couverture de la fibre peuvent être endogènes aux dynamiques économiques locales. Néanmoins, ce problème est vraisemblablement moins prononcé dans les zones d'initiative publique dont relève l'essentiel de notre échantillon d'estimation.

Des effets positifs sur les dynamiques de création d'établissements.

- Nos résultats en régression de panel montrent une association positive entre la hausse du taux de couverture fibre et la variation nette du stock d'établissements ainsi que des créations brutes de nouveaux établissements.
- Les effets sont toutefois d'une ampleur limitée : passer d'une couverture de 0 à 100% est ainsi associé à une hausse d'environ 2 % du nombre d'établissements

au niveau communal. Les régressions au niveau secteur–année suggèrent des élasticités comparables.

- L'effet sur l'emploi est positif mais non significativement différent de zéro dans l'échantillon principal. On constate une baisse du nombre de demandeurs d'emploi d'environ -1 %.
- Ces résultats sont globalement compatibles avec l'idée que l'extension de la fibre a permis de créer des emplois via la création d'établissements ayant peu de salariés.
- Les résultats obtenus sur le sous-échantillon des communes en RIP sont qualitativement proches de ceux obtenus sur l'ensemble des échantillons.
- Les résultats statiques sont globalement confirmés par les études d'évènements qui ne détectent pas de pré-tendances différentielles antérieures au traitement.

Des corrélations positives sur la valeur des mutations, mais des tendances pré-traitement différentielles.

- Nos résultats en régression de panel montrent une association positive entre la hausse du taux de couverture fibre et la variation du nombre de mutations et des valeurs des transactions.
- Passer d'une couverture de 0 à 100% est associé à une hausse d'environ +0.4 % du nombre de mutations annuelles au niveau communal.
- Les résultats sur la valeur des transactions suggèrent une absence d'effet sur la **valeur médiane** mais un effet positif sur la **valeur moyenne** (+2,7 %).
- Cependant, les études d'évènements détectent, notamment pour la valeur totale des transactions, des tendances différentielles antérieures au traitement. L'interprétation causale des coefficients statiques n'est donc pas soutenue par l'analyse dynamique.

Chapitre 2 : Accessibilité de la fibre et résilience au choc sanitaire

- Ce chapitre vise à comprendre l'effet du déploiement de la fibre sur la dynamique du marché du travail et de l'immobilier lors de la **crise liée à l'épidémie de Covid-19** à partir du deuxième trimestre de 2020.
- Concernant le marché du travail, nous utilisons des données à fréquence trimestrielle sur les offres d'emplois postées ainsi que sur les embauches. Ces données, disponibles jusqu'à la mi-2021, permettent de capter les effets de la crise sur la dynamique du marché du travail de façon beaucoup plus fine que les données d'emploi dans la mesure où le recours massif à l'activité partielle a grandement stabilisé le niveau de l'emploi.
- Notre approche principale consiste à comparer le taux de croissance des variables d'intérêt au sein des communes couvertes par la fibre à plus de 75 % au 2020T1 à celui au sein des communes non-couvertes.
- Afin de nous concentrer sur un ensemble de communes comparables, nous restreignons notre échantillon d'estimation à des communes où le déploiement de la fibre était nul au début de l'année 2019 (2019T1). Dès lors, le statut de couverture début 2020 dépend uniquement du progrès du déploiement au cours de l'année 2019.
- Nous montrons que cet échantillon de communes est largement constitué ($\approx 85\%$) de communes rurales, de faible densité et relevant des Réseaux d'initiatives publiques (RIP). Il est donc particulièrement pertinent du point de vue de l'évaluation du Plan France Très Haut Débit.

Le marché du travail des communes connectées à la fibre au début de 2020 a mieux résisté à la crise.

- L'échantillon a subi un choc massif avec un déclin de 50 % des offres postées au 2020T2. Les communes couvertes par le déploiement de la fibre entre

2019 et 2020 ont néanmoins relativement mieux résisté à la crise : le taux de croissance des offres est environ 10 points de pourcentage plus élevé qu’au sein des communes non-couvertes de notre échantillon. Cet écart est également perceptible au niveau des embauches avec un écart d’environ 8 points de pourcentage.

- Bien que le statut de traitement ne soit naturellement pas alloué de façon aléatoire, **plusieurs éléments viennent soutenir la plausibilité d’une interprétation causale des écarts statistiquement significatifs documentés** entre communes couvertes et non-couvertes. En premier lieu, nous montrons que le coefficient associé à la couverture est très peu sensible à l’introduction de nombreuses variables de contrôle mesurant des caractéristiques observables des communes : composition sectorielle de l’emploi, caractère rural ou de montagne, densité. En second lieu, nous montrons que les communes couvertes et non-couvertes suivaient des tendances largement parallèles entre 2019T2 et 2018T2 et également entre 2020T1 et 2019T1. L’hypothèse d’identification de cette approche est que les communes “traitées” auraient suivi une tendance similaire à celle des communes du groupe contrôle en l’absence de couverture. Bien que non-directement testable, cette hypothèse est confortée par le fait qu’avant le choc Covid et la couverture fibre, les deux groupes de communes suivaient des tendances parallèles.

Le marché de l’immobilier a également mieux résisté.

- Au niveau du **marché immobilier**, nous documentons une forte baisse agrégée entre 2020 et 2019 du nombre de transactions. Là encore, cette baisse est moins marquée, d’environ 18 %, au sein des communes couvertes par la fibre.
- Notre approche est robuste à la constitution d’un groupe de contrôle alternatif qui ne contient que les villes ayant une couverture nulle au 2019T1 mais **ayant connu une avancée des travaux** entre 2019T1 et 2020T1.

L'adoption du télétravail apparaît comme un mécanisme naturel.

- Au niveau **individuel**, nous documentons une adoption plus forte du télétravail en 2020 dans les occupations intensives en numérique au sein des communes connectées par rapport aux communes non-connectées à la fibre.

Chapitre 3 : Analyse des effets du déploiement de la fibre au niveau infra-communal

- Ce chapitre s'attache à analyser à un niveau spatial très fin (la parcelle cadastrale), les effets d'une connexion à la fibre (mesurée comme le raccordement d'un point de mutualisation) sur des dimensions comme la population vivant dans la parcelle, les revenus de ses habitants, et le marché immobilier local.
- L'analyse au niveau de la parcelle cadastrale permet à la fois d'exploiter des variations très fines au niveau spatial en comparant des zones traitées et non-traitées plus proches que dans le cas de l'approche communale des chapitre 1 et 2.
- Elle permet également d'obtenir une taille d'échantillon largement supérieure (environ 25M d'observations).
- Nos résultats indiquent un impact positif, que l'on peut interpréter comme étant causal, de la connexion au THD, sur la démographie des parcelles cadastrales traitées, qui voient le nombre de leurs foyers fiscaux augmenter (+0,1 %) et l'âge moyen des adultes diminuer ($\approx -0,25$ année d'âge moyen). Les effets au sein des RIP sur ces variables sont également significatifs, plus élevés pour le nombre de foyers (+0,3 %) mais plus faibles pour l'âge moyen (-0,1 année).
- Le raccordement au THD a également un impact positif sur la probabilité de revenus industriels et commerciaux (+0,1 pp dans l'échantillon total, +0,15

pp au sein des RIP). On constate également un impact négatif sur la probabilité de percevoir des pensions de retraite.

- On ne détecte cependant pas d'effet robuste sur les revenus moyens du fait d'une violation très probable de l'hypothèse de tendance commune suggérée par des tendances pré-traitement très différentes. Ces pré-tendances suggèrent globalement que les ménages vivant au sein de parcelles traitées plus précocement avaient tendance à voir leurs revenus croître plus rapidement.
- Ces résultats sur la structure des revenus et sur la structure démographique sont d'ailleurs en cohérence, et suggèrent une évolution du profil des zones connectées à la fibre vers une population plus jeune et vers une activité non-salariée plus dynamique.
- Ceci invite naturellement à examiner les dynamiques immobilières, qui pourraient en théorie refléter une telle évolution démographique. Les résultats sont plus mitigés.
- Le taux de vacance des logements diminue dans les parcelles raccordées, mais seulement au sein d'un sous-échantillon spécifique. De plus, la taille des effets est très modeste : -0,2 point de pourcentage.
- On constate une hausse de la probabilité qu'une entreprise soit implantée dans les parcelles cadastrales raccordées. Néanmoins, on remarque des pré-tendances différentielles. Cela suggère que les zones les plus dynamiques aient été connectées en premier, même à une échelle spatiale très fine.
- Nous ne constatons pas d'effet robuste sur les prix immobiliers pratiqués au sein des parcelles raccordées. Ce résultat peut être rationalisé vis-à-vis de la littérature existante par des rendements décroissants de la vitesse des accès internet fixes.

SOMMAIRE

Introduction et synthèse des résultats	15
1 Analyse des effets du déploiement de la fibre au niveau communal	19
1.1 Données	19
1.1.1 Problèmes liés aux données par classes de débit	19
1.1.2 Approche par type de technologie	23
1.1.3 Principales bases de données et variables d'intérêt	25
1.2 Statistiques Descriptives	27
1.3 Approche Empirique	32
1.3.1 Régression de panel au niveau commune et commune-secteur	33
1.3.2 Étude d'évènement autour de l'évènement de la hausse	34
1.4 Résultats : emploi, démographie des entreprises et marché immobilier	35
1.4.1 Emploi et démographie des entreprises : analyse dynamique.	35
1.4.2 Emploi et démographie des entreprises : analyse statique au niveau communal.	42
1.4.3 Emploi et démographie des entreprises : dimension commune-secteur	48
1.5 Résultats : marché Immobilier	49
1.6 Conclusion	54
1.7 Approche dynamique : évolution du taux de couverture.	55
1.8 Le choix de l'échantillon d'estimation pour les études d'évènements .	57
2 Accessibilité de la fibre et résilience à la crise sanitaire	61
2.1 Introduction	61
2.2 Données	63
2.3 Stratégie empirique	65
2.4 Statistiques descriptives	66
2.5 Résultats principaux	70
2.5.1 Dynamique des offres d'emploi et des embauches (2020T2) .	70
2.5.2 Marché immobilier	76
2.5.3 Tests de robustesse	77

2.6	Résultats complémentaires : télétravail et activité partielle sur données individuelles	78
2.6.1	Données individuelles	79
2.6.2	Spécification empirique	80
2.6.3	Résultats sur télétravail	81
2.7	Conclusion	84
2.8	Annexe	85
2.8.1	Construction de l'index de télétravaillabilité : L'enquête Conditions de Travail 2019	86
2.8.2	Tests de robustesse	90
3	Analyse des effets du déploiement de la fibre au niveau local	95
3.1	Données	95
3.2	Approche empirique	97
3.2.1	Étude d'évènement échelonnée	97
3.2.2	Étude d'évènement empilée avec sélection explicite du groupe de contrôle	99
3.3	Impact de la connexion au THD sur la population et les revenus . . .	105
3.4	Impact de la connexion au THD sur le marché immobilier local . . .	116
3.5	Conclusion	129
	Annexe	131
A.1	Discussion des hypothèses d'identification de l'étude d'événements . .	131
	Références	135

CHAPITRE 1

ANALYSE DES EFFETS DU DÉPLOIEMENT DE LA FIBRE AU NIVEAU COMMUNAL

Ce premier chapitre présente nos analyses et résultats au niveau communal concernant l'évolution de la couverture Très Haut Débit (THD) sur la période 2015-2020, et ses effets sur la démographie des entreprises et le marché immobilier. La section 1.1 présente les données utilisées et souligne notamment les problèmes liés aux données par classes de débit. Nous présentons ensuite des statistiques descriptives sur le déploiement de la fibre entre 2015 et 2020 (le coeur de notre période d'évaluation) dans la section 1.2. Ensuite, la section 1.3 explicite notre approche empirique. La section 1.4 présente les principaux résultats.

1.1 Données

1.1.1 Problèmes liés aux données par classes de débit

L'existence d'une multiplicité de sources de données non-concordantes impose de définir avec précision notre objet d'intérêt. Par souci de transparence méthodologique, nous détaillons dans cette partie les étapes du raisonnement qui nous ont poussé à retenir le taux de couverture communal en termes d'accès à la fibre

comme variable d'intérêt.

Les données de l'Agence du numérique sont disponibles de 2013 à 2018T1 et contiennent un taux de couverture communal pour chaque classe de débit descendant. L'utilisation de ces données permet donc *a priori* de s'intéresser avec finesse à l'effet de l'accès à un certain niveau de débit, c'est-à-dire à une certaine qualité de l'accès à Internet, indépendamment des technologies mises en place pour obtenir ce niveau de débit, ce qui est un objectif pertinent pour l'évaluation du Plan France Très Haut Débit.

Cependant, l'étendue temporelle de ces données est limitée, et il a semblé particulièrement important d'étendre l'analyse aux années 2018 et 2019, dans la mesure où ce se sont celles qui ont vu une forte hausse du taux de couverture en France avec la montée en charge du Plan France Très Haut Débit (PFTHD).

Cependant, une telle extension des données de l'Agence du numérique s'est révélée problématique, car les données candidates utilisées pour une telle extension ne permettent pas de reconstituer un panel homogène permettant de mesurer des taux de couverture THD cohérents (défini sur la base d'une classe de débit) au niveau communal. Ces données supplémentaires sont issues du site ma-connexion, et présentent des taux de couverture toutes technologies confondues par classes de débit, permettant ainsi d'isoler la part d'une commune éligible au THD.

Deux limites principales s'opposent à l'utilisation de ces données. La première est qu'elles ne sont disponibles qu'à partir du T3 2019, alors même que les données de l'Agence du numérique s'arrêtent au T1 2018, ce qui impose une forme d'interpolation ou plus généralement d'imputation pour la période intermédiaire. La deuxième consiste en l'existence de fortes divergences de taux entre les deux sources (ma-connexion et Agence du numérique). Le tableau 1.1 résume les distributions de taux de couverture THD au niveau communal entre ces deux sources.

Le taux de couverture moyen à l'échelle du pays, c'est-à-dire toutes communes confondues, est plus élevé dans les données issues de l'Agence du numérique (au

T1 2018) que dans les données ma-connexion, pourtant plus tardives (T3 2019). La différence est substantielle (8 points de pourcentage) et soulève un doute sur la qualité des données. Cette différence ne tient pas à une sur-représentation de valeurs manquantes ou de taux de couverture nuls dans la base de données ma-connexion ; ainsi le nombre de taux nuls est inférieur au T3 2019 (données ma-connexion) par rapport au T1 2018 (données Agence du numérique), comme l'indique le Tableau 1.2.

TABLEAU 1.1 – Distribution du taux de couverture THD, données Agence du numérique (T1 2018) vs données ma-connexion (T3 2019)

	N obs.	moyenne	p25	p50	p75	p90	p95
Taux de couverture THD (T1 2018)	35342	0,26	0,00	0,01	0,51	0,85	0,95
Taux de couverture THD (T3 2019)	35005	0,18	0,00	0,01	0,31	0,59	0,78

Note : Données Agence du numérique T1 2018 vs ma-connexion T3-2019. Comparaison des taux de couverture THD et distribution

TABLEAU 1.2 – Décompte - communes non éligible au THD , données Agence du numérique (T1 2018) vs données ma-connexion (T3 2019)

	T1 2018	T3 2019
Taux nuls	16858	15857

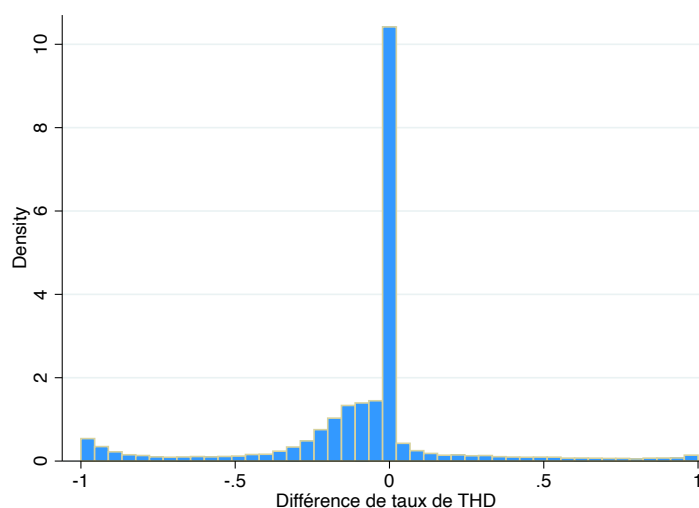
Note : Données Agence du numérique T1 2018 vs ma-connexion T3-2019. Comptage des communes où le taux de couverture est égal à 0

La figure 1.1 présente la distribution de la différence au niveau de chaque commune présente dans les deux bases entre le taux présent dans ma-connexion (T3 2019) et le taux de l'Agence du numérique (T1 2018). Il est théoriquement possible que cette différence prenne une valeur négative, en particulier du fait de la construction de locaux non-connectés. En effet, la différence de taux de couverture peut être formellement définie de la manière suivante :

$$\text{Différence}_{\text{Commune C}} = \frac{\# \text{ locaux connectés}_{\text{Commune C, T2}}}{\# \text{ locaux}_{\text{Commune C, T2}}} - \frac{\# \text{ locaux connectés}_{\text{Commune C, T1}}}{\# \text{ locaux}_{\text{Commune C, T1}}}$$

Ainsi, à déploiement constant, un projet immobilier qui augmente substantiellement le nombre de locaux dans une commune ($\# \text{ locaux}_{\text{Commune } C, T2}$) entraîne une différence négative si ces locaux ne sont pas immédiatement connectés après leur apparition dans les données cadastrales. Cependant, le nombre de municipalités pour lesquelles le taux de couverture THD est plus faible dans ma-connexion que dans les données Agence du numérique est très important ; de plus, le coefficient de corrélation entre ces deux mesures est assez bas (0,47).

FIGURE 1.1 – Différence de taux de couverture THD pour une commune donnée,



Note : Évolution des taux de couverture THD pour une commune donnée, entre les données ma-connexion (T3 2019) et les données Agence du numérique (T1 2018). On s'attendrait à ce qu'une grande majorité des différences soient positives en principe.

L'ensemble de ces éléments nous ont conduit à abandonner l'exploitation de la source ma-connexion. Nous adoptons à la place une approche par types de technologie, qui permet de rendre compte d'une partie seulement du déploiement des technologies numériques dans le cadre du PFTHD, mais avec plus de précision et avec des données cohérentes.

1.1.2 Approche par type de technologie

L'ARCEP met à disposition des fichiers de taux de couverture par type de technologie et par commune du T4 2018 au T2 2021¹. La principale limite dans l'utilisation de ces données est qu'elles ne contiennent que des taux de couverture par type de technologie et non par classe de débit. Ceci signifie qu'il est impossible de reconstruire un taux de couverture correspondant au concept de "très haut débit", utilisé dans les données de l'Agence du numérique dans la période précédente (débit descendant supérieur à 30 Mb/s). Les données de l'Agence du numérique contiennent également, à partir de 2015, des informations sur le taux de couverture par la technologie de la fibre ftth (tel que disponible dans les données de l'Arcep).

Le principal désavantage d'une telle redéfinition (taux de couverture fibre plutôt que taux de couverture très haut débit) est qu'elle ne permet donc pas d'appréhender dans son ensemble le plan France Très Haut Débit. Elle constitue donc une limite à l'exhaustivité de notre démarche empirique. Néanmoins, le déploiement de la fibre constitue le coeur du PFTHD et fournit donc la source de variation la plus pertinente, étant données les sources disponibles, du point de vue de l'évaluation du PFTHD.

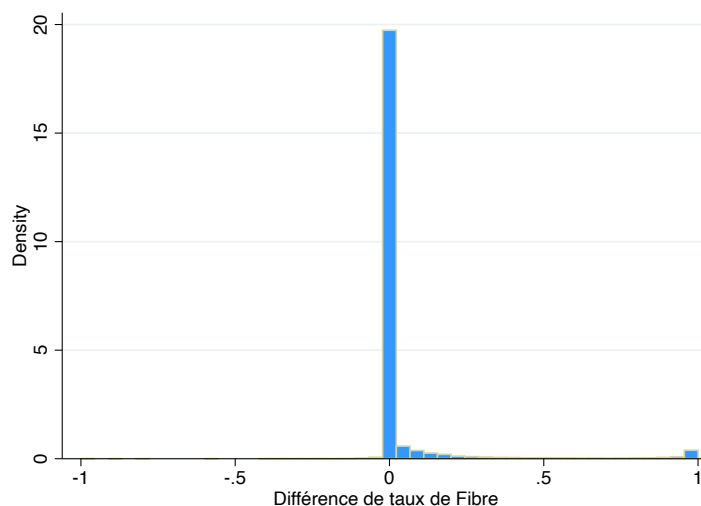
Par souci de transparence, nous réalisons le même type de comparaisons entre les deux sources pour en vérifier la cohérence et la qualité. De manière générale, les données fournies par l'Arcep ont l'avantage de présenter une grande cohérence en termes de taux de couverture (par technologie et non pas débit) avec celles fournies par l'Agence du numérique pour la période début 2015-début 2018.

En effet, on constate que les différences de taux de couverture au sein d'une commune donnée entre les deux sources sont très limitées, ce qui suggère que les séries sont largement comparables. La Figure 1.2 présente les différences au sein

1. Les données sont disponibles en ligne : <https://www.data.gouv.fr/fr/datasets/le-marche-du-haut-et-tres-haut-debit-fixe-deploiements/>.

des communes entre d'une part la source Agence du numérique au 2018 T1 et d'autre part la source ARCEP au 2019 T1.

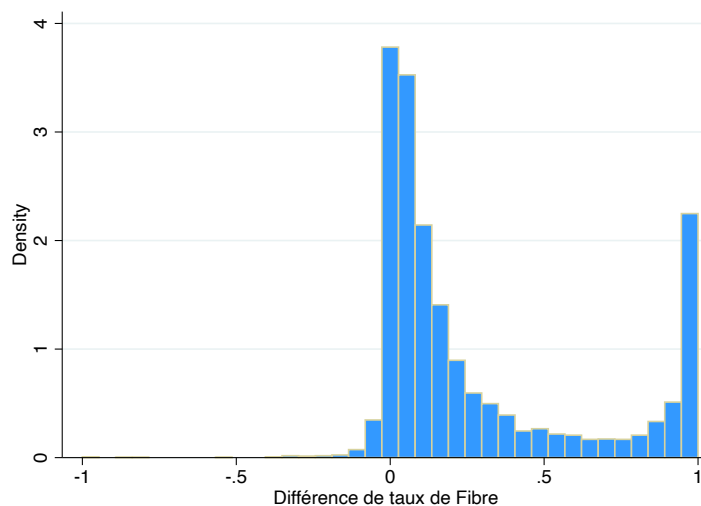
FIGURE 1.2 – Différence de taux de couverture Fibre pour une commune donnée,



Note : Évolution des taux de couverture Fibre pour une commune donnée, entre les données fibre ARCEP open data (T4 2018) et les données ARCEP précédentes (T1 2018). On s'attendrait à ce que toutes les différences soient positifs en principe.

La figure 1.3 montre qu'une très grande majorité des taux reste inchangés entre les deux trimestres considérés. Les divergences de taux sont en très grande majorité positives, avec parmi les quelques cas négatifs, des taux très proches de 0. Ces évolutions à la baisse du taux de couverture semblent donc marginales, et compatibles avec des évolutions réelles du taux de couverture (par exemple sous l'effet d'une augmentation du nombre total de bâtiments plus rapide que celle du nombre de bâtiments connectés).

FIGURE 1.3 – Différence de taux de couverture Fibre pour une commune donnée, en excluant les années sans variation



Note : Evolution des taux de couverture Fibre pour une commune donnée, entre les données fibre ARCEP open data (T4 2018) et les données Agence du numérique précédentes (T1 2018) On s'attendrait à ce que toutes les différences soient positifs en principe

1.1.3 Principales bases de données et variables d'intérêt

Répertoire des entreprises et des établissements (REE). Nous documentons la démographie des entreprises grâce aux bases de stocks issues du Répertoire des Entreprises et des Etablissements (REE/Sirene) produit par l'Insee. Les bases que nous utilisons regroupent les établissements actifs en France métropolitaine au 31 décembre de l'année considérée. Nous considérons le champ marchand. Nous utilisons cette base afin de construire le stock d'établissement (nombre d'établissements actifs au 31/12) ainsi que les créations d'établissements (nombre de créations au cours de l'année t rapporté au nombre d'établissements actifs au 31/12/ $t - 1$).

Immobilier. La Direction Générale des Finances Publiques (DGFIP) a développé et rendu public le fichier "Demandes de Valeurs Foncières" (DVF) qui recense l'ensemble des mutations foncières à titre onéreux publiées dans les services de la pu-

blicité foncière. Cette base de données fiscales est structurée, vérifiée et enrichie par le Cerema qui produit la base de données DV3F, qui contient des données foncières complémentaires issues des fichiers fonciers.

Dans ce chapitre, nous utilisons cette base afin de calculer le nombre de mutations par commune et année.

Emploi. La structure de l'emploi au niveau communal est issue des fichiers de l'Acoss. L'Acoss met en ligne les dénombrements des établissements employeurs et des effectifs salariés, par catégorie d'activité (nomenclature NAF à 732 postes) et par commune, au 31 décembre. Nous utilisons ces données afin d'évaluer l'évolution de l'emploi.

Recensement. De la même manière, nous complétons la description de la structure des économies locales par plusieurs fichiers issus du recensement. Tout d'abord, les caractéristiques socio-démographiques (CSP, part de retraités) en 2017 proviennent du recensement de l'Insee. D'autres caractéristiques des communes ont également été extraites des bases de données de l'Insee, comme la catégorie de densité de population, le nombre de ménages, le nombre de logements, le degré d'urbanisation de la commune et son caractère montagnoux. Ces informations sont évaluées entre 2013 et 2015 dans le cadre de ce chapitre, pour rendre compte des caractéristiques des communes avant le déploiement du PFTHD.

Variables d'intérêts. Afin de saisir tous les effets potentiels d'une plus grande couverture fibre, nous nous focalisons principalement sur trois catégories de variables d'intérêt : celles relatives à l'emploi, celles relatives au dynamisme économique et à la création d'entreprises, et celles relatives au marché immobilier local. Nous mobilisons pour ces variables d'intérêt plusieurs bases de données différentes, et en particulier : les données Acoss pour l'emploi salarié par commune,

le Répertoire des Entreprises et Etablissements (REE) pour mesurer la création et le stock d'établissements, le recensement de la population établi par l'Insee afin de saisir l'évolution de la structure socio-économique (structure par catégories socio-professionnelles, nombre de chômeurs, catégorie d'âge, etc.) de chaque commune ainsi que l'évolution de sa population, les données IRCOM (DGFIP et Insee) pour l'évolution des revenus des ménages et la base de données DV3F (DGFIP et Ceremma) contenant des informations sur les valeurs foncières et le nombre de mutations.

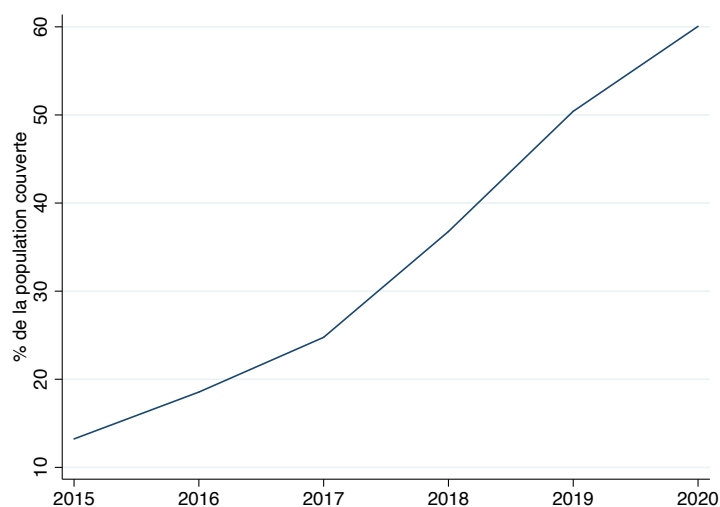
1.2 Statistiques Descriptives

Taux de couverture fibre. Cette section présente les principales tendances observables quant à l'évolution communale de la couverture fibre. Le taux de couverture au niveau communal est défini comme la part des locaux situés au sein de la commune qui sont connectés à la fibre. La couverture fibre totale a fortement augmenté depuis 2015. Comme le montre la figure 1.4, fin 2020 presque 60 % de la population métropolitaine avait un accès à la fibre, contre seulement un peu plus de 10 % début 2015.

Le tableau 1.3 présente la distribution de trois statistiques : (a) la différence de taux de couverture entre deux années pour une même commune, (b) la différence totale entre 2015 et 2020 pour une même commune et (c) la différence entre deux années lors de l'année de la plus forte augmentation. Un nombre important d'observations² correspondent à des communes observées pendant une année où le taux de couverture n'a pas évolué (à peu près la moitié). Environ 10 % des observations correspondent à des augmentations annuelles communales du taux de couverture d'au moins 76 points de couverture. Si 25 % des communes n'ont vu leur taux de couverture augmenter que de 50 points, 25 % des communes sont complètement

2. Une observation correspond ici à une commune observée une année donnée.

FIGURE 1.4 – Part de la population couverte



Note : Données ARCEP 2015-2020, taux de couverture agrégé par commune pondérés par la population communale. Le taux de couverture est défini comme la part des locaux d'une commune qui sont éligibles à la fibre.

couvertes en 2020, avec un taux avoisinant les 100 % de couverture.

Enfin, il est intéressant de noter que (c) est très proche de (b) ce qui indiquerait que l'évolution de la couverture ne fut pas progressive. Beaucoup de communes voient leur taux de couverture augmenter de façon ponctuelle et importante, avec une augmentation très importante une année donnée, plutôt qu'une augmentation plus étalée dans le temps. 25 % des communes sont ainsi passées d'un taux de couverture presque nul à une couverture complète, quand en moyenne l'année de la plus forte progression le taux de couverture fibre a augmenté de 64 points. Ainsi on peut dire que l'évolution de la couverture fibre entre 2015 et 2020 correspond, en première approximation, à un schéma d'extension commune par commune.

Le tableau 1.4 confirme cette observation et classe les années de chaque commune en trois groupes : celle de la plus forte hausse, celle(s) antérieure(s) à cette année là et celle(s) postérieure(s). On remarque ainsi que, dans la majorité des cas, la hausse est concentrée lors d'une année donnée, le taux de couverture n'évoluant

TABLEAU 1.3 – Évolution brute du taux de couverture par ville

	Moyenne	écart-type	p25	p50	p75	p90	p95
Différence annuelle (a)	0,15	0,30	0,00	0,00	0,13	0,76	0,97
Différence moyenne totale (b)	0,72	0,32	0,51	0,87	0,98	1,00	1,00
Meilleure année (c)	0,64	0,33	0,36	0,72	0,96	1,00	1,00

Note : Données ARCEP 2015-2020. Distribution du taux de couverture fibre 2015-2020 par commune. a) représente la différence annuelle (entre deux années), b) la différence totale entre 2015 et 2020 pour une même commune et c) l'évolution du taux lors de la meilleure année/année de la plus forte augmentation. Le taux de couverture est défini comme la part des locaux d'une commune qui sont éligibles à la fibre.

peu suite à cette hausse soudaine. En effet, celui-ci n'augmente en moyenne que de 4 points maximum par an par la suite, même si pour 10 % des communes la hausse peut atteindre 14 points par an.

TABLEAU 1.4 – Catégorisation des années : année de la hausse la plus importante, postérieure et antérieure

	N obs.	moyenne	écart-type	p25	p50	p75	p90	p95
Année(s) antérieure(s)	66278	0,01	0,04	0,00	0,00	0,00	0,05	0,10
Année de la hausse	13268	0,67	0,30	0,41	0,76	0,97	1,00	1,00
Année(s) postérieure(s)	14067	0,04	0,08	0,00	0,01	0,07	0,14	0,20
Observations	93613							

Note : Données ARCEP 2015-2020. Répartition des observations en fonction des années et de leur groupe. Chaque observation correspond à une année x commune. Chaque observation est ensuite répartie en fonction de si elle correspond à 2) l'année de la plus forte hausse, 3) une année postérieure à la plus forte hausse 1) une année antérieure. Le taux de couverture est défini comme la part des locaux d'une commune qui sont éligibles à la fibre.

Ce type d'extension très soudain est utile car il nous permettra de mettre en place une estimation par "étude d'évènement" (Sun et Abraham, 2021; Borusyak et al., 2021)—voir section 1.3.2.

Le tableau 1.5 présente pour chaque année le nombre de communes ayant connu une croissance significative de leur taux de couverture.³ On constate qu'au-

3. Par construction la première année de l'échantillon (2015) pour laquelle on observe le taux de

tour de 14 000 ont connu une hausse de leur taux de couverture fibre, tandis que 20 500 communes n'ont pas connu de hausse significative au cours de la période.

On observe de cette façon une intensification de la croissance de la couverture fibre, avec une part importante de villes ayant connu leur plus forte augmentation en 2019 et 2020.

TABLEAU 1.5 – Catégorisation des années

	2015	2016	2017	2018	2019	2020
Année de la hausse	0	671	934	2335	4064	5965
Année(s) postérieure(s)	0	0	668	1605	3940	8004
Année(s) antérieure(s)	13969	13298	12367	10029	5965	0
Pas de hausse	20488	20488	20488	20488	20488	20488
Total	34457	34457	34457	34457	34457	34457

Note : Données Agence du numérique-ARCEP 2015-2020. Répartition des observations en fonction des années et de leur groupe. Chaque observation correspond à une année x commune. Chaque observation est ensuite répartie en fonction de si elle correspond à 1) l'année de la plus forte hausse, 2) une année postérieure à la plus forte hausse 3) une année antérieure. Si la commune n'a jamais connu de hausse significative de son taux de couverture entre 2015 et 2020 toutes ses observations vont dans le groupe 4) pas de hausse. Par construction il n'y a pas d'observations en 2015 pour 1), en 2015 et 2015 pour 2), et en 2020 pour 3).

TABLEAU 1.6 – Distribution des taux de couverture, 2015-2020

	N obs.	moyenne	écart-type	p25	p50	p75	p90	p95
Taux Fibre en 2015	34457	0,01	0,10	0,00	0,00	0,00	0,00	0,00
Taux Fibre en 2020	34457	0,31	0,42	0,00	0,00	0,81	0,98	1,00
Taux Fibre en 2015, non nuls	1091	0,47	0,32	0,17	0,44	0,77	0,90	0,94
Taux Fibre en 2020, non nuls	14440	0,73	0,32	0,55	0,88	0,98	1,00	1,00

Note : Données ARCEP 2015-2020. Distribution des taux de couverture fibre en 2015 et 2020

Statistiques descriptives sur l'ensemble des communes et les cohortes 2016 à 2020. Nous présentons des statistiques descriptives sur le niveau des variables principales en 2015 (début de notre échantillon d'observations) et leur évolution moyenne annuelle entre 2015 et 2019. Nous distinguons deux échantillons. Le pre-

couverture de la fibre ne peut pas être l'année de la plus forte augmentation car on observe pas le taux de couverture fibre en 2014. De la même façon 2020 ne peut pas être une année "antérieure" à celle ci et 2015/2016 ne peuvent pas être des années postérieures.

mier comprend l'ensemble des communes pour lesquelles nous sommes en mesures de suivre l'activité économique (emploi, établissement) au cours de la période 2015 à 2019. Le second comprend uniquement les villes qui ont connu une forte hausse entre 2015 et 2020 et qui n'étaient pas fortement couvertes dès 2015 (moins de 25% de taux de couverture).

TABLEAU 1.7 – Statistiques descriptives : ensemble des communes

	Observations	Moy.	écart-type	p25	p50	p75
Variables en niveau : moyenne 2013-2015						
Densité h/km ²	34457	163,429	732,729	18,850	41,008	96,681
Population	34457	1830,143	15196,054	200	452	1128
Part sans diplôme	34457	0,354	0,111	0,279	0,347	0,422
Part éducation supérieur	34457	0,201	0,097	0,136	0,190	0,253
Nombre établissements	34457	162,160	2733,701	20	39	90
Emploi salarié	34457	501,519	8229,577	3	18	112
Évolution moyenne 2015-2019						
Δ%stock	34457	0,033	0,044	0,010	0,031	0,052
Δ%étab, empl,	30612	0,001	0,099	-0,028	0,000	0,029
Δ%étab, empl,	30080	0,006	0,155	-0,030	0,000	0,046
Δ%valeur fonc, moy,	32659	0,029	0,403	-0,090	0,026	0,145
Δ%valeur fonc, tot,	32609	0,060	0,640	-0,139	0,066	0,253
Taux de création	34457	0,072	0,037	0,046	0,069	0,095
Catégorie commune						
		part	Cohortes	Nb com.	part	
Part de communes RIP	34457	0,827	Cohorte 2016	671	0,048	
Part de communes ZTD	34457	0,003	Cohorte 2017	934	0,067	
Part de communes AMI	34457	0,102	Cohorte 2018	2,335	0,167	
Part de communes AMEL	34457	0,053	Cohorte 2019	4,064	0,291	
Com. non-couvertes en 2020	34457	0,57	Cohorte 2020	5,965	0,427	
Com. couvertes en 2015	34457	0,02				

Notes : Données ARCEP 2015-2020. Statistiques descriptives pour l'ensemble des communes de l'échantillon apparié. La densité est mesurée en habitants au kilomètre carré. La variable de population est mesurée en nombre d'habitants. Le symbole Δ% fait référence à la variation en pourcent (taux de croissance). Les variables de densité, population, part des habitants adultes sans diplômes et avec diplôme de l'enseignement supérieure sont évaluées à leur moyenne entre 2013 et 2015.

TABLEAU 1.8 – Statistiques descriptives : cohortes 2016-2020

	Observations	Moy.	écart-type	p25	p50	p75
Variables en niveau : moyenne 2013-2015						
Densité, h/km2	13969	321,206	1124,500	33,939	75,967	195,736
Population	13969	3530,330	23729,696	315	778	2185
Part sans diplôme	13969	0,338	0,101	0,268	0,331	0,400
Part éducation supérieure	13969	0,217	0,096	0,151	0,206	0,270
Nombre établissements	13969	310,766	4287,551	25	56	165
Emploi salarié	13969	1095,623	12896,265	7	48	363
Demandeurs d'emploi	13969	319,261	2329,314	20	50	155
Évolution moyenne 2015-2019						
Δ%stock	13969	0,038	0,042	0,017	0,036	0,056
Δ%étab, empl,	13028	0,003	0,086	-0,020	0,000	0,028
Δ%étab, empl,	12895	0,009	0,136	-0,021	0,006	0,040
Δ%valeur fonc, moy,	12677	0,033	0,341	-0,072	0,026	0,125
Δ%valeur fonc, tot,	12666	0,068	0,497	-0,103	0,074	0,232
Taux de création	13969	0,082	0,037	0,057	0,081	0,106
Catégorie commune				Nb communes	part	
RIP	13969	0,743	Cohorte 2016	671	0,048	
ZTD	13969	0,008	Cohorte 2017	934	0,067	
AMII	13969	0,229	Cohorte 2018	2,335	0,167	
AMEL	13969	0,008	Cohorte 2019	4,064	0,291	
			Cohorte 2020	5,965	0,427	

Notes : Données ARCEP 2015-2020. Statistiques descriptives pour l'ensemble des communes de l'échantillon ayant connu une hausse significative de leur taux de couverture entre 2015 et 2020. La densité est mesurée en habitants au kilomètre carré. La variable de population est mesurée en nombre d'habitants. Le symbole Δ% fait référence à la variation en pourcent (taux de croissance). Les variables de densité, population, part des habitants adultes sans diplômes et avec diplôme de l'enseignement supérieure sont évaluées à leur moyenne entre 2013 et 2015. La densité est exprimée en habitants au kilomètre carré, la population en habitants.

1.3 Approche Empirique

Afin d'analyser les conséquences du plan France Très Haut Débit, nous proposons deux approches empiriques, la première statique, la deuxième dynamique.

1.3.1 Régression de panel au niveau commune et commune-secteur

L'approche statique consiste à comparer l'évolution de nos variables d'intérêts, au niveau commune (marché immobilier) ou commune x secteur (création d'établissements, nombre d'établissements, emploi salarié ...), entre fonction de l'évolution du taux de couverture. Exploitant des données riches de panel, au format commune × année, on peut définir le modèle suivant pour chaque commune i de l'année t .

$$Y_{it} = \beta \times D_{it} + \alpha_i + \lambda_t + \epsilon_{it} \quad (1.1)$$

où Y_{it} est la valeur prise par notre variable d'intérêt, pour la commune i et l'année t . D_{it} correspond au taux de couverture fibre, α_i est un vecteur d'effets fixes au niveau commune, λ_t un vecteur d'effets fixes au niveau année, et ϵ_{it} est un terme d'erreur.

La spécification concernant les variables au niveau commune × secteur est quelque peu différente et est décrite par le modèle suivant pour chaque secteur s de la commune i et de l'année t :

$$Y_{its} = \beta \times D_{its} + \alpha_{is} + \lambda_t + \epsilon_{its} \quad (1.2)$$

où Y_{its} vaut pour chacune de nos variables d'intérêt, α_{is} est un vecteur d'effets fixes au niveau commune - secteur, λ_t un vecteur d'effets fixes au niveau année, et ϵ_{its} est un terme d'erreur. On intègre des effets fixes années, communes et secteurs, et on clusterise au niveau commune⁴.

4. Une troisième spécification est envisagée afin de mieux prendre en compte l'effet sur les créations d'entreprises. En effet il est important de prendre en compte qu'à la dimension secteurs NAF 17 x année x commune, toutes les communes n'ont pas nécessairement observées une création d'établissements dans chaque secteur. Utiliser le logarithme du nombre de créations peut donc biaiser les résultats avec beaucoup de zéros. De cette façon il est possible d'envisager une régression suivant le modèle de Poisson. La spécification est donc la suivante :

$$y_{ist} = \exp(x_{ist} \times \beta + \alpha_{is} + \lambda_t) + \epsilon_{ist}$$

1.3.2 Étude d'évènement autour de l'évènement de la hausse

L'approche dynamique vise à estimer la dynamique des variables d'intérêt au sein des communes autour des années de forte progression du taux de couverture de celles-ci. Cette approche exploite précisément le fait que ces épisodes de forte progression n'ont pas eu lieu au même moment pour chaque commune. Chaque commune qui a vu son taux progresser se voit attribuer une année t , dite de la plus forte augmentation. On calcule ensuite pour chaque année de notre échantillon, l'écart entre celle-ci et t . Par exemple si $t = 2018$ pour la commune A, $t - 1 = 2017$, $t + 1 = 2019$, etc. Nous estimons le modèle suivant :

$$y_{it} = \alpha_i + \lambda_t + \sum_{d \neq -1} \beta_d \times \mathbb{1}\{d = t - t_i^0\} + \varepsilon_{it} \quad (1.3)$$

pour chaque variable d'intérêt y défini pour la commune i dans l'année t . Le terme t_i^0 fait référence à l'année de traitement pour la commune i . Les paramètres α_i et λ_t représentent des effets fixes communes et villes, permettant de prendre en compte les particularités de chaque commune et de chaque année. Afin d'isoler l'effet de l'année t_i^0 vis à vis des années précédentes, on normalise en fixant l'année $t_i^0 - 1$ ainsi que la première année de l'échantillon à 0, ce qui normalise les années de pré-traitement (années antérieures à l'année t_i^0). Notre échantillon est également sélectionné de façon à ce que, peu importe l'année t , chaque commune possède le même nombre d'années de pré-traitement. Dans la section A.1 de l'annexe, nous discutons les hypothèses d'identification qui sous-tendent cette approche plus en détails.

Où y_{ist} est la probabilité qu'une création d'un secteur s dans une commune i et une année t arrive, b_{it} étant le coefficient de régression du taux de fibre, α_i et λ_t des effets fixes villes et années, et ε_{ist} un terme d'erreur. Les erreurs sont également clusterisés au niveau commune.

Le choix de l'échantillon d'estimation. Cette approche se concentre sur les communes ayant connu une forte hausse de leur taux de couverture en 2016, 2017, 2018, 2019 et 2020. La période d'étude comprend les années 2011 à 2019. La cohorte de communes traitées en 2020 sert donc de contrôle dans cette analyse. La Section 1.8 en annexe explique les raisons motivant ces restrictions. Ce choix est principalement justifié par le fait que les pré-tendances sur des variables clés (comme le stock d'établissements) sont très différentes, ce qui rend peu plausible la validité de l'hypothèse de tendances communes en l'absence de déploiement du THD entre les communes traitées entre 2016 et 2020 et celles qui le seront plus tard.

Les deux sections suivantes présente nos résultats empiriques sur la démographie des entreprises (section 1.4) ainsi que sur le marché de l'immobilier (section 1.5).

1.4 Résultats : emploi, démographie des entreprises et marché immobilier

1.4.1 Emploi et démographie des entreprises : analyse dynamique.

Nous présentons ici les résultats de l'analyse dynamique (étude d'évènements ou double différence dynamique). Les figures présentent deux séries de coefficients et leurs intervalles de confiance à 95 %. La première série "TWFE" correspond à l'estimation par moindres carrés ordinaires de l'équation dynamique (voir équation 1.3) en incluant les effets fixes années et villes. La série dite "Sun-Abraham" repose sur un estimateur alternatif développé par Sun et Abraham (2021) qui a l'avantage de ne pas être biaisé en présence d'hétérogénéité des effets entre cohortes, où la co-

horte est définie comme l'ensemble des communes recevant le traitement la même année.

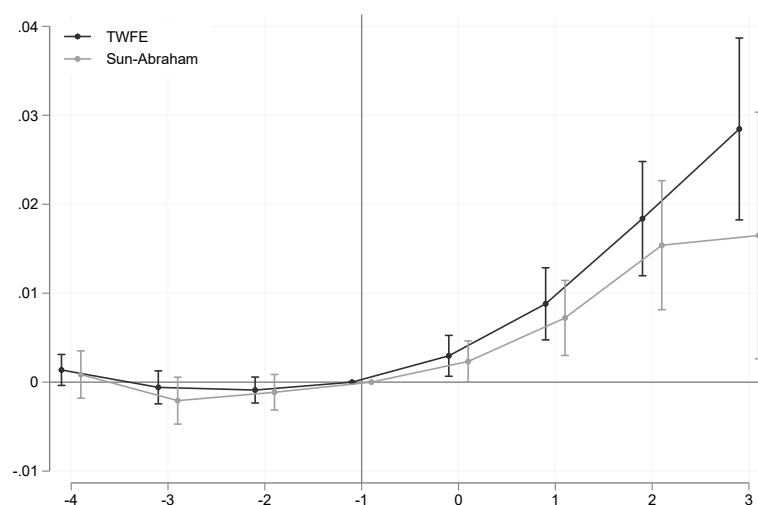
En annexe, nous présentons en outre des éléments sur la dynamique du taux de couverture de la fibre qui supporte la faisabilité de cette approche, qui repose sur une date de traitement discrète bien que le taux de couverture fibre soit une variable continue (section 1.7). Nous présentons également plus en détails au sein de l'annexe la nature de l'échantillon retenu (section 1.8). Nous montrons les résultats dynamiques sur le stock d'établissements (figure 1.5) et sur le nombre brut de nouvelles créations d'établissements (figure 1.6).

La figure 1.5 suggère un important effet sur le nombre d'établissement avec un coefficient aux alentours de +2 % à la fois selon la méthode conventionnelle et la méthode de Sun-Abraham (ci-après SA). Dans les deux cas, on ne détecte pas de pré-tendances différentielles très élevées (même si elles sont parfois significativement différentes de 0). On constate dans l'ensemble des résultats assez similaires, avec des effets moyens positifs.

La Figure 1.6) présente les résultats sur le nombre de créations d'établissements. On constate un effet positif, bien que peu précisément estimé. Cette variable est un flux et tend à être plus dispersée ce qui explique peut-être la moindre précision par rapport aux résultats sur le stock dans le graphique précédent. Au vu des résultats dans les figures 1.5 et 1.6, la présence d'un effet positif sur la démographie des établissements semble très plausible. Nous présenterons plus de détails sur la nature des établissements potentiellement créés dans l'analyse statique, notamment concernant leur appartenance sectorielle.

Avant de présenter les résultats statiques, nous présentons les résultats sur l'emploi. La figure 1.7 montre qu'il est difficile de mettre en évidence un effet positif sur l'emploi. L'estimateur SA est relativement imprécis, en revanche l'estimateur conventionnel peut rejeter des effets de +2 % avec un niveau de confiance de 95 % sur l'ensemble de la période post-traitement. La figure 1.8 s'intéresse au nombre

FIGURE 1.5 – Effets Dynamiques - Nombre d'établissements



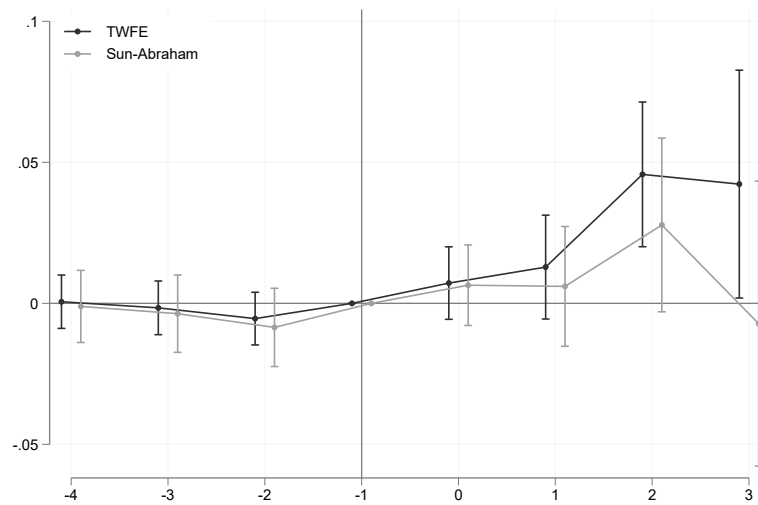
Note : DiD Dynamique : effets des années pré et post hausse sur le log du nombre d'établissements.2015-2019. Effets fixes année + commune et clustérisés au niveau commune. Effets normalisés à t-1 et t-5. Intervalles de confiance à 95 %. Les effets sont calculés de sorte que peu importe l'année de la hausse,le nombre d'années de pré-tendance soit le même.

d'établissements employeurs et confirme la difficulté de détecter des effets sur les variables liées à l'emploi salarié dans le cadre de l'approche dynamique.

Enfin, la figure 1.9 présente les résultats sur le nombre de demandeurs d'emploi en fin de mois. Les résultats sont moins clairs que pour les autres variables. On constate des coefficients post-traitement largement négatifs autour de -1 % pour les période $d = +2, +3$. Néanmoins, cet effet semble en partie transitoire et se rapproche de 0 à $d = +3$. De plus, la méthode SA qui trouve les effets les plus nets semble également suggérer que la réduction dans le nombre de demandeurs d'emploi est intervenus en partie avant la hausse du taux de couverture.

Le fait que que l'on trouve un effet négatif sur le nombre de demandeurs d'emploi sans trouver d'effet sur le nombre de salariés pourrait être expliqué par le fait que l'accès à la fibre a permis à des activités non-salariées d'émerger au sein des communes traitées—ce qui est reflété par la hausse du stock et des créations d'éta-

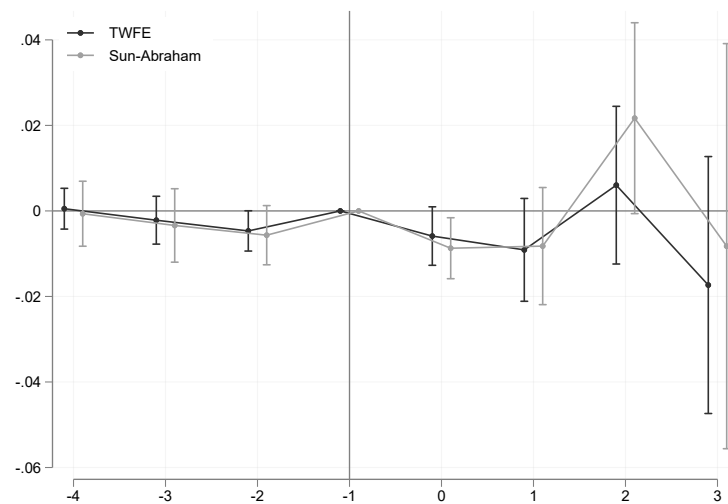
FIGURE 1.6 – Effets Dynamiques - Nombre de créations



Note : Double Différence Dynamique : effets des années pré et post-hausse sur le log du nombre de créations +1. 2015-2020 pondéré par le stock l'année précédent l'année de la hausse. Intervalles de confiance à 95 %. Effets fixes année, commune et clustérisés au niveau commune. Normalisés à t-1 et t-5 Les effets sont calculés de sorte que peu importe l'année de la hausse, le nombre d'années de pré-tendance soit le même

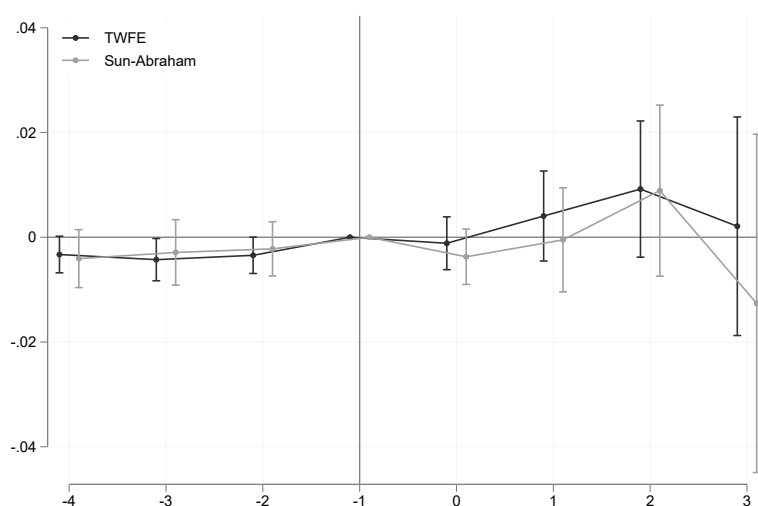
blissements.

FIGURE 1.7 – effets dynamiques - nombre de salariés



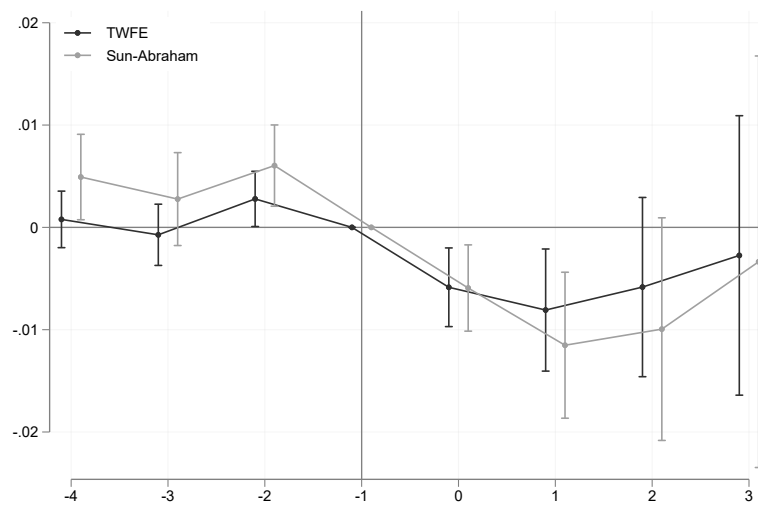
notes : double différence dynamique : effets des années pré et post-hausse sur le log du nombre de salariés pondérés par le stock de l'année précédent l'année de la hausse. les effets sont calculés de sorte que peu importe l'année de la hausse, le nombre d'années de pré-tendance soit le même. intervalles de confiance à 95 %. effets fixes année et commune. écarts-types clustérisés au niveau commune. coefficients normalisés à $t - 1$ et $t - 5$.

FIGURE 1.8 – effets dynamiques - nombre d'établissements employeurs



notes : double différence dynamique : effets des années pré et post-hausse sur le log du nombre de salariés pondérés par le stock de l'année précédent l'année de la hausse. les effets sont calculés de sorte que peu importe l'année de la hausse, le nombre d'années de pré-tendance soit le même. intervalles de confiance à 95 %. effets fixes année et commune. écarts-types clustérisés au niveau commune. coefficients normalisés à $t - 1$ et $t - 5$.

FIGURE 1.9 – effets dynamiques - nombre de demandeurs d'emploi



Notes : double différence dynamique : effets des années pré et post-hausse sur le log du nombre de salariés pondérés par le stock de l'année précédent l'année de la hausse. les effets sont calculés de sorte que peu importe l'année de la hausse, le nombre d'années de pré-tendance soit le même. intervalles de confiance à 95 %. effets fixes année et commune. écarts-types clustérisés au niveau commune. coefficients normalisés à $t - 1$ et $t - 5$.

1.4.2 Emploi et démographie des entreprises : analyse statique au niveau communal.

Analyse sur l'ensemble des secteurs pour l'échantillon complet. Nous nous intéressons aux effets de la fibre sur la démographie des entreprises via l'estimation du modèle statique.

Le tableau 1.9 présente les résultats de l'estimation de l'équation (1.1). L'échantillon d'estimation inclut l'ensemble des communes françaises n'ayant pas fait l'objet d'une fusion ou bien d'une division lors de la période 2015-2020. Il inclut donc à la fois les villes ayant connu une hausse de la couverture fibre avant 2015, les villes ayant connu un déploiement fibre entre 2016 et 2020 et celles n'ayant pas encore été connectées en 2020.

On constate un effet positif sur le nombre d'établissement (colonne 1). Passer de 0 à 1 en termes de taux de couverture (c'est-à-dire passer d'une couverture inexistante à une couverture totale) implique une hausse de 2 % du nombre d'établissements au niveau communal. Cette hausse est économiquement importante : faire passer le taux de couverture passe de 0 à 100 % implique une hausse de l'ordre de 40 % de l'écart entre le premier et le dernier quartile en terme de croissance annuelle du nombre d'établissement entre 2015 et 2019 (voir Table 1.7). La colonne (2) présente l'impact sur le (log du) nombre d'établissements employeurs. L'effet est positif mais deux fois plus faible que l'effet sur le stock total d'établissements. Cet écart implique que la part des établissements employeurs baisse suite au déploiement de la fibre. On constate dans la colonne (3) un effet positif, faible et marginalement significatif sur le nombre de salariés. Cet effet est inférieur à celui sur le nombre d'établissements (colonne 1), ce qui implique que la taille moyenne des établissements a tendance à baisser. Le déploiement de la fibre semble donc associé à une hausse relativement plus importante de l'activité non-salariée.

On constate une hausse du nombre de créations d'établissements (colonne 4)

avec un effet dont l'ampleur est globalement similaire à l'effet sur le stock d'établissements. Il apparaît donc que l'effet sur le stock d'établissements reflète en grande partie, au sein de cet échantillon, un effet sur l'entrée de nouveaux établissements plutôt qu'un effet sur la persistance d'établissements existants (moins de cessation d'activité ou mobilité par exemple).

Enfin, on constate un effet négatif sur le nombre de demandeurs d'emploi. Le coefficient estimé présenté dans la colonne (5) implique que passer de 0 à 1 est associé à une baisse d'environ 0,6 % du nombre moyen de demandeurs d'emploi en fin de mois. Une hausse du taux de couverture de fibre de 0 à 100 % correspond à 4,7 % de l'écart inter-quartiles du taux de croissance du nombre de demandeurs d'emplois. Cette baisse est donc significative statistiquement mais modeste au vu de la dispersion de l'évolution du nombre de demandeurs d'emplois entre communes de l'échantillon.

TABLEAU 1.9 – Emploi et démographie des entreprises

	(1) # étab.	(2) # étab. empl.	(3) emploi (salarié)	(4) # création	(5) # chômeurs
Taux couverture fibre	0,0202*** (0,00165)	0,00746* (0,00383)	0,00344 (0,00529)	0,0247*** (0,00669)	-0,0104*** (0,00265)
R^2	0,996	0,989	0,989	0,908	0,991
Effets fixes com.	✓	✓	✓	✓	✓
# communes	34457	30727	30233	34457	30850
Observations	172267	149515	146780	172267	148980

Note : Ce tableau présente les coefficients d'une étude d'évènements ou différence-de-différences dynamiques (voir équation 1.1). Les écart-types robustes clustérisés au niveau commune sont reportés entre parenthèses. La variable # étab. (colonne 1) fait référence au log du stock d'établissement (source : REE). La variable # étab. empl. est le log du stock d'établissement employeur (source ACOSS). La variable emploi (salarié) est le log de l'emploi salarié (source ACOSS). La variation # création (colonne 4) est le log du nombre de créations + 1 (source REE). Enfin, la colonne (5) présente le log du nombre moyen par an de demandeurs d'emploi en fin de mois (source INSEE/Pôle emploi). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

L'échantillon principal contient environ 83 % de communes RIP (voir le panel inférieur "Catégorie commune" du Tableau 1.7). Les effets estimés jusqu'à pré-

sent sont donc susceptibles de refléter en grande partie l'influence de ces communes. Afin de nous assurer que les résultats sont stables au sein des communes RIP nous effectuons l'estimation statique sur ces communes uniquement. Les résultats sont présentés dans le tableau 1.10. Les résultats sur le stock d'établissements et le nombre de demandeurs d'emploi sont similaires aux effets estimés sur la base de l'échantillon entier. Le coefficient concernant la création d'emploi est positif mais n'est pas significativement différent de 0 bien qu'il ne soit pas statistiquement différent de l'effet sur le stock d'établissement. Les résultats confirment des effets proches de 0 concernant l'emploi et les établissements employeurs.

TABLEAU 1.10 – Emploi et démographie des entreprises : échantillon RIP

	(1) # étab.	(2) # étab. empl.	(3) emploi (salarié)	(4) # création	(5) # chômeurs
Taux couverture fibre	0,0117*** (0,00184)	0,00287 (0,00446)	-0,00476 (0,00619)	0,00762 (0,00762)	-0,0155*** (0,00305)
R^2	0,995	0,985	0,986	0,876	0,987
Effets fixes com.	✓	✓	✓	✓	✓
# communes	28479	25108	24668	28479	25236
Observations	142377	121891	119463	142377	121498

Note : Ce tableau présente les coefficients d'une étude d'évènements ou différence-de-différences dynamiques (voir équation 1.1). Les écart-types robustes clustérisés au niveau commune sont reportés entre parenthèses. La variable # étab. (colonne 1) fait référence au log du stock d'établissement (source : REE). La variable # étab. empl. est le log du stock d'établissement employeur (source ACOSS). La variable emploi (salarié) est le log de l'emploi salarié (source ACOSS). La variation # création (colonne 4) est le log du nombre de créations + 1 (source REE). Enfin, la colonne (5) présente le log du nombre moyen par an de demandeurs d'emploi en fin de mois (source INSEE/Pôle emploi). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Analyse sur un groupe restreint de secteurs intensifs en usage numérique.

Des travaux précédents (Mack et Rey, 2014) suggèrent que la qualité de la connexion internet est un facteur de choix de localisation des entreprises plus important parmi les secteurs intensifs en savoirs et utilisations numériques. Il est donc plausible que l'effet du déploiement de la fibre ait un impact sur l'emploi et le dynamisme écono-

miques qui soient concentrés ou a minima plus marqués au sein de ces secteurs.

Ainsi, nous proposons deux nouvelles régressions sur un échantillon réduit à deux ensembles de secteurs : d'une part, les secteurs a priori intensifs en usages numériques (Information et communications, activités financières et d'assurance, activités immobilières, activités scientifiques et techniques ; services de soutien) et d'autre part le secteur de l'hôtellerie-restauration.⁵ Dans les deux cas, on distingue l'effet sur l'échantillon total et celui sur l'échantillon RIP.

Le tableau 1.11 présente les résultats de l'estimation de l'équation (1.1) pour les secteurs intensifs en numérique. On constate un effet positif sur le nombre d'établissements (colonne 1) ainsi que le nombre d'établissements employeurs (colonne 2). L'effet sur l'emploi salarié (colonne 3) est plus marqué que pour l'ensemble des secteurs (+2,4 % par rapport à +0,3 %) et est significatif. On voit enfin un effet très net sur les créations d'établissements (+3,9 %). Lorsque l'on se focalise sur les communes RIP dans le tableau 1.12, on constate des résultats assez proches concernant le stock d'établissements et l'emploi salarié. En revanche, on ne trouve d'effet ni sur les créations ni sur les établissement employeurs.

Globalement, en dépit des différences entre les deux échantillons, on constate globalement des effets plus marqués sur cet ensemble de secteurs.

5. Nous remercions Joëlle Toledano pour cette suggestion.

TABLEAU 1.11 – Emploi et démographie des entreprises - secteurs intensifs en numérique

	(1) # étab.	(2) # étab. empl.	(3) emploi (salarié)	(4) # création
Taux couverture fibre	0.0144*** (0.00374)	0.0144*** (0.00349)	0.0241*** (0.00585)	0.0388*** (0.00656)
R^2	0.984	0.984	0.982	0.872
Effets fixes com.	✓	✓	✓	✓
# communes	34457	34457	34457	34457
Observations	172267	172267	172267	172267

Note : Erreurs standard entre parenthèses. Échantillon dimension commune x année, limité à certains secteurs naf A17. Effets Fixes Communes + Années ; Cluster niveau Commune. Nombre de communes : 34457

TABLEAU 1.12 – Emploi et démographie des entreprises - secteurs intensifs en numérique au sein des communes RIP

	(1) # étab.	(2) # étab. empl.	(3) emploi (salarié)	(4) # création
Taux couverture fibre	0,00822* (0,00429)	0,00573 (0,00390)	0,0149** (0,00644)	0,00759 (0,00723)
R^2	0,978	0,977	0,976	0,816
Effets fixes com.	✓	✓	✓	✓
# communes	28479	28479	28479	28479
Observations	142377	142377	142377	142377

Note : Erreurs standard entre parenthèses. Échantillon dimension commune x année, limité à certains secteurs naf A17. Effets Fixes Communes + Années ; Cluster niveau Commune. Nombre de communes : 34457

Concernant le secteur hôtellerie-restauration, les résultats présentés dans les tableaux 1.13 et 1.14 sont nuancés. Au sein de l'échantillon complet, on voit des effets plus faibles sur les établissements et les établissements employeurs mais un effet plus positif pour le nombre brut de création d'établissement ainsi que l'emploi.

Ce résultat est compatible avec un processus de destruction créatrice où le taux de création brut a augmenté et a induit la sortie de certains établissements plus petits, de telle sorte que le nombre d'établissements n'évolue pas et l'emploi salarié augmente. En revanche, nous ne sommes pas en mesure de détecter un effet significatif pour le sous-échantillon des communes RIP sur nos variables d'intérêt.

TABLEAU 1.13 – Emploi et démographie des entreprises - hôtellerie-restauration

	(1) # étab.	(2) # étab. empl.	(3) emploi (salarié)	(4) # création
Taux couverture fibre	0,00559* (0,00332)	0,00536* (0,00297)	0,0195*** (0,00508)	0,0214*** (0,00481)
R^2	0,981	0,979	0,978	0,781
Effets fixes com.	✓	✓	✓	✓
# communes	34457	34457	34457	34457
Observations	172267	172267	172267	172267

Note : Erreurs standard entre parenthèses. Échantillon dimension commune x année, limité à certains secteurs naf A17. Effets Fixes Communes + Années; Cluster niveau Commune. Nombre de communes : 34457

TABLEAU 1.14 – Emploi et démographie des entreprises - hôtellerie-restauration au sein des communes RIP

	(1) # étab.	(2) # étab. empl.	(3) emploi (salarié)	(4) # création
Taux couverture fibre	-0,00572 (0,00368)	-0,000806 (0,00321)	0,00536 (0,00549)	0,0000404 (0,00481)
R^2	0,976	0,972	0,971	0,708
Effets fixes com.	✓	✓	✓	✓
# communes	28479	28479	28479	28479
Observations	142377	142377	142377	142377

Note : Erreurs standard entre parenthèses. Échantillon dimension commune x année, limité à certains secteurs naf A17. Effets Fixes Communes + Années; Cluster niveau Commune. Nombre de communes : 34457

1.4.3 Emploi et démographie des entreprises : dimension commune-secteur

Dans cette sous-section, nous procédons à une estimation au niveau secteur \times commune. Cette approche s'intéresse aux effets de la fibre au sein d'un couple commune-secteur donné. Elle permet de répondre aux biais potentiels liés par exemple à la composition sectorielle locale. En regardant l'effet au niveau secteur-commune, on s'intéresse à l'évolution de la démographie des entreprises au sein d'un secteur donné avant et après le déploiement de la fibre dans la commune.

Le tableau 1.15 présente les résultats de l'estimation de l'équation (1.2). On ne constate pas d'effet sur le taux de création d'établissement (colonne 1). Cette variable étant définie comme le ratio du nombre de création une année donnée par rapport à la valeur du stock d'établissements l'année précédente, il est possible que cette régression confonde des effets à la fois sur le numérateur et de le dénominateur qui se compensent. On considère donc le logarithme du nombre d'établissement dans la colonne (2) et on constate un effet positif et significatif mais modeste : passer de 0 à 1 en termes de taux de couverture implique une hausse de 1,8 % du nombre d'établissement. La colonne (3) présente le nombre de créations qui augmenté d'environ 1,3. Cet effet positif est confirmé par les colonnes (5) et (6) qui considèrent le log du nombre de créations + 1 et présente une régression de type pseudo-Poisson respectivement. Enfin, la colonne (4) présente les résultats de notre estimation sur le nombre de salariés et trouve un effet positif.

Les résultats sont globalement comparables lorsque des effets fixes année - secteurs sont inclus (Tableau 1.16), ce qui suggère que les différences de compositions sectorielles sont peu susceptibles d'expliquer les effets estimés au niveau commune sur l'activité économique.

TABLEAU 1.15 – Taux de couverture Fibre et activité économique

	(1) # étab.	(2) # étab. empl.	(3) emploi (salarié)	(4) # création	(5) Créat. > 0
Taux couverture fibre	0,0154*** (0,00158)	0,00966*** (0,00178)	0,0101*** (0,00286)	0,0448*** (0,00219)	0,00511*** (0,00180)
R^2	0,978	0,980	0,979	0,817	0,537
Effets fixes com. × secteur	✓	✓	✓	✓	✓
Année × secteur	✓	✓	✓	✓	✓
# communes	34457	30659	30152	34457	34457
Observations	1704958	833847	797839	1785468	1785468

Note : Ce tableau présente les résultats de l'estimation de l'équation (1.2). Erreurs standard entre parenthèses. Effets Fixes Communes x Secteurs + Années; Cluster niveau Commune. Nombre de communes : 34457.

TABLEAU 1.16 – Taux de couverture Fibre et activité économique : effet fixes secteur × année.

	(1) # étab,	(2) # étab. empl.	(3) emploi (salarié)	(4) # création	(5) # chômeurs
Taux couverture fibre	0,0147*** (0,00158)	0,0107*** (0,00179)	0,0100*** (0,00287)	0,0444*** (0,00217)	-0,00773*** (0,00235)
R^2	0,978	0,980	0,979	0,816	0,993
Effets fixes com. × secteur	✓	✓	✓	✓	✓
# communes	34457	30659	30152	34457	30850
Observations	1704958	833847	797839	1785468	1654388

Note : Ce tableau présente les résultats de l'estimation de l'équation (1.2). Erreurs standard entre parenthèses. Effets Fixes Communes x Secteurs + Années x Secteurs; Cluster niveau Commune. Nombre de communes : 34457.

1.5 Résultats : marché Immobilier

En plus de regarder les effets sur l'activité économique, nous proposons de regarder l'effet de la fibre sur le marché immobilier local.

Les caractéristiques des logements sont valorisées par les acheteurs ou locataires potentiels et affectent les prix de l'immobilier ainsi que les loyers (Rosen, 1974). Ces caractéristiques peuvent être propres au logement (nombre de pièces par exemple)

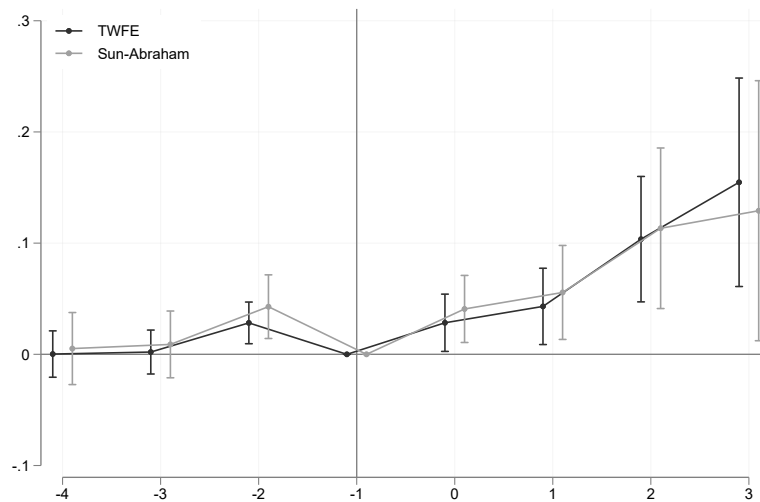
ou bien dépendre de son environnement plus ou moins immédiat du fait de sa localisation (qualité des écoles par exemple). L'accès potentiel à une connexion internet THD fait également partie de l'ensemble des caractéristiques des logements susceptibles d'être valorisées⁶.

La fibre étant un service supplémentaire, elle apporte donc potentiellement une plus-value à un bien et peut apporter un plus grand intérêt immobilier à une commune. Par exemple, un individu télé-travaillant trouvera peut-être davantage intérêt à emménager dans une commune bien connectée et aura dès lors une disposition à payer plus élevée. On pourrait donc attendre un effet positif de la fibre sur les prix et le nombre de transactions effectuées localement.

Analyse dynamique. Concernant l'analyse dynamique des effets de la fibre sur l'immobilier, on s'intéresse dans un premier lieu à l'impact de la couverture sur les valeurs foncières moyennes des communes (Figure 1.10). On constate un effet positif, quelque soit la méthode d'estimation retenue, de l'ordre de +10 %. Bien que les pré-tendances soient relativement plates, l'ampleur de ces effets semble trop grande pour refléter un effet pur de valorisation de la fibre. Il est possible que ce coefficient reflète des effets de composition dans les biens qui font l'objet d'une transaction du fait de la fibre. La figure 1.11 considère la valeur totale des transactions ayant lieu au sein d'une commune, c'est-à-dire le (log du) produit de la valeur moyenne et du nombre de transactions. Les résultats suggèrent un effet très important +75 % dans le cas de l'estimation TWFE. Néanmoins, deux éléments nuancent ce constat : les tendances pré-traitement sont différentes et l'estimation SA aboutit à des coefficients beaucoup plus modestes de l'ordre de +10 %.

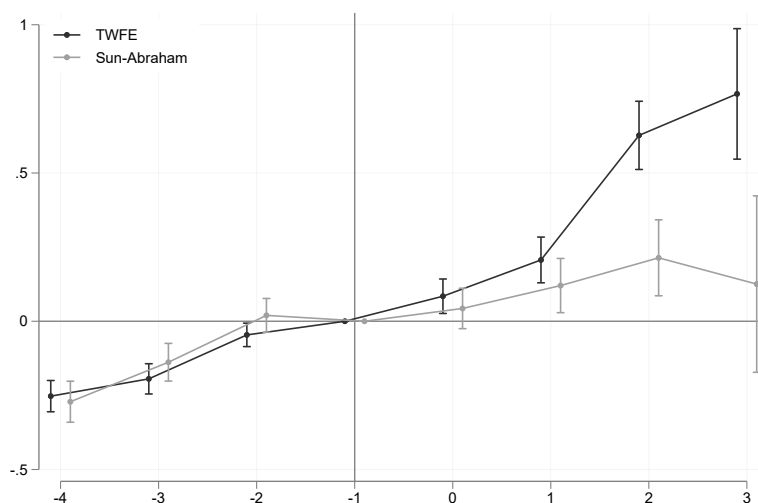
6. Par exemple, (Chay et Greenstone, 2005) analyse empiriquement les prix de l'immobilier au niveau des comtés américains afin de mesurer la capitalisation des régulations environnementales dans le prix de l'immobilier, reposant ainsi sur la théorie des prix hédoniques appliquée au marché immobilier.

FIGURE 1.10 – Effets Dynamiques - Valeur foncière moyenne



Note : Double Différence Dynamique : effets des années pré et post hausse sur le log de la valeur foncière moyenne. 2015-2019 pondérés par le stock l'année précédent l'année de la hausse. Effets normalisés à t-1 et t-5. Effets fixes année, commune et clustérisés au niveau commune. Les effets sont calculés de sorte que peu importe l'année de la hausse, le nombre d'années de pré-tendance soit le même. Intervalles de confiance à 95 %

FIGURE 1.11 – Effets Dynamiques - Valeur totale des transactions



Note : Double Différence Dynamique : effets des années pre et post hausse sur le log du nombre de transactions. 2015-2019 pondérés par le stock l'année précédent l'année de la hausse. Effets normalisés à $t - 1$ et $t - 5$. Effets fixes année, commune et clustérisés au niveau commune. Les effets sont calculés de sorte que peu importe l'année de la hausse, le nombre d'années de pré-tendance soit le même. Intervalles de confiance à 95 %

Analyse statique. Le tableau 1.17 présente les résultats des régressions statiques sur une variable égale à 1 si une transaction a eu lieu au sein de la commune considérée, le log de la valeur totale des transactions immobilières, le log de la valeur moyenne de cette transaction et le log de la valeur moyenne.

On constate colonne (1) un effet positif sur la probabilité qu'un bien immobilier dans la commune fasse l'objet d'une transaction. La valeur totale (colonne 2), qui reflète à la fois le nombre et la valeur moyenne des transactions, augmente de +8 %. La valeur moyenne augmente de +2,7 % (colonne 3). Cet effet est plus petit pour la valeur médiane et il n'est pas significativement différent de 0 (colonne 4). Il est possible que seuls certains segments du marché immobilier soient sensibles à l'accès à la fibre ce qui expliquerait pourquoi la valeur moyenne augmente sans que la valeur médiane ne réagisse dans les mêmes proportions. Néanmoins, dans les deux cas les coefficients ne sont pas très précisément estimés et ne sont pas statistiquement différents l'un de l'autre.

TABLEAU 1.17 – Taux de couverture Fibre et Marché Immobilier

	(1) Transaction	(2) Valeur totale	(3) Valeur moy.	(4) Valeur méd.
Taux couverture fibre	0,00361* (0,00202)	0,0888*** (0,0326)	0,0274** (0,0138)	0,0154 (0,0179)
R^2	0,769	0,812	0,534	0,469
Effets fixes com.	✓	✓	✓	✓
# communes	34457	34457	32766	32766
Observations	172267	172267	160053	160053

Note : Erreurs standard entre parenthèses. Effets Fixes Communes + Années; Cluster niveau Commune. Nombre de communes : 34457

La hausse de la valeur moyenne pourrait refléter à la fois une hausse globale de la demande d'immobilier pour la commune du fait d'effets d'entraînement et la hausse de la demande liée spécifiquement aux biens immobiliers disposant d'un accès à la fibre. L'accès à des données plus fine (niveau infra-communal) permettra

de déterminer si cette hausse est bien concentrée ou non sur les biens disposant effectivement d'un accès potentiel à la fibre.

1.6 Conclusion

Ce chapitre s'est intéressé aux effets du déploiement de la fibre entre 2015 et 2019 au sein des communes françaises. La période est caractérisée par un déploiement important. Une partie importante de ce déploiement a lieu par communes, c'est-à-dire que pour beaucoup de communes, il est possible d'identifier une année clef, qui représente l'année principale du déploiement, du point de vue de la magnitude de l'augmentation du taux de couverture des locaux de la commune. Ceci justifie la construction d'une étude d'évènement autour de la date du déploiement principal de chaque commune, en complément d'une approche statique plus classique.

Le déploiement de la fibre a des effets notables sur la démographie des entreprises d'une commune. Le nombre des établissements au niveau communal progresse de manière économiquement et statistiquement significative avec la couverture fibre d'une commune. Passer de 0% à 100% de couverture fibre est associé à une augmentation de 2% du nombre d'établissements. Il s'agit avant tout d'un effet portant sur la création de nouveaux établissements. Des effets négatifs, bien que modestes, sur le nombre de chômeurs, une absence d'effet clairement identifiable sur l'emploi salarié, et une magnitude plus faible des effets sur le nombre d'établissements employeurs : tout porte à croire que ce dynamisme de la démographie des entreprises passe avant tout par l'emploi non-salarié et la création de petits établissements.

Les effets sont plus marqués statistiquement et économiquement sur les secteurs intenses en technologie numériques, ce qui est attendu par la théorie économique ; l'emploi salarié dans ces secteurs est également stimulé par le taux de couverture

fibre, ce qui est à comparer à l'effet total nul sur l'emploi salarié en général. Les effets dans l'hôtellerie-restauration sont plus ambigus et semblent pointer vers un effet positif du déploiement, et une concentration vers les plus gros établissements. De manière générale, les résultats sont qualitativement similaires dans les communes RIP que dans les autres communes, mais d'une ampleur économiquement moins significative.

Nous cherchons à trouver trace de ces résultats sur le marché immobilier, qui pourrait refléter les effets positifs sur la démographie des entreprises. La valeur foncière moyenne dans les communes semble réagir positivement à l'augmentation du taux de couverture par la fibre, sans qu'il soit facile d'interpréter cet effet comme la résultante directe de la valorisation de l'accès à la fibre pour les locaux connectés, plutôt que comme un effet d'entraînement plus général sur l'attractivité globale de la commune.

Annexe

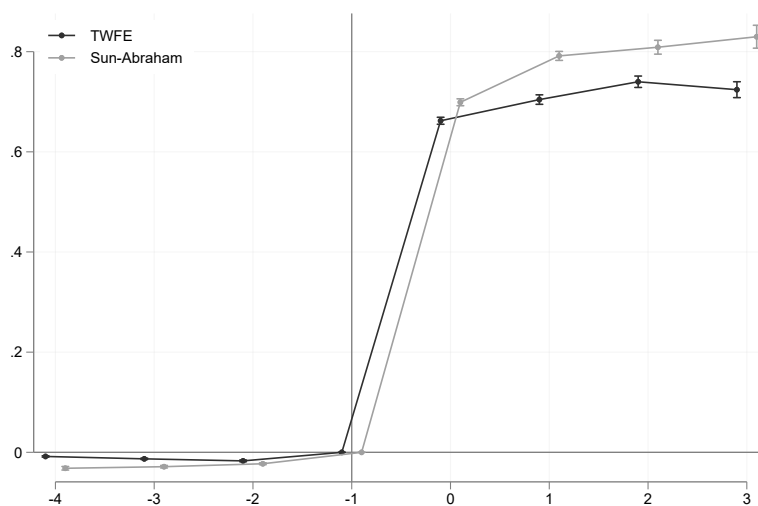
1.7 Approche dynamique : évolution du taux de couverture.

Nous implémentons une approche dynamique qui permet à la fois d'estimer les effets en fonction du temps écoulé depuis le traitement et d'estimer les pré-tendances. Le taux de couverture est une variable continue, néanmoins comme expliqué ci-dessus (voir Tableau 1.3), les communes tendent à voir leur taux augmenter de façon très soudaine au cours d'une année donnée qui correspond à la hausse de la variable continue la plus élevée. On peut dès lors attribuer cette année comme date de traitement et mettre en place une étude d'évènement en comparant les dynamiques avant et après cette date.

La figure 1.12 vérifie en premier lieu l'évolution du taux de couverture fibre autour de l'année de la plus forte augmentation (année t). De façon assez attendue, par définition, on remarque un effet moyen positif de l'année de la plus forte augmentation sur le taux de couverture fibre. On remarque cependant que la hausse est concentrée dans les deux années suivant cette date, de telle sorte que la discrétisation de la variable continue de couverture afin de déterminer une date ne semble pas causer une grande perte d'information.

Les deux estimateurs donnent des résultats extrêmement proches : on constate que la hausse du taux de couverture est concentrée autour de la date d'augmentation principale et passe de 0 à 80 %. Dans le reste de l'analyse nous comparons la spécification standard (estimé par OLS avec effets fixes ville et année) avec l'approche de Sun et Abraham (2021).

FIGURE 1.12 – Effets Dynamique - Taux de couverture



Note : Cette figure présente les coefficients d'une étude d'évènements ou différence-de-différences dynamiques (voir équation 1.3). Les intervalles de confiance à 95% sont représentés sur la base d'écart-types robustes clustérisés au niveau commune. La variable dépendante est le taux couverture fibre, défini comme la part des locaux d'une commune éligibles à la fibre.

1.8 Le choix de l'échantillon d'estimation pour les études d'évènements

La mise en place d'études d'évènements rend la question du choix du groupe de contrôle particulièrement saillante.

Compte tenu des données choisies, nous disposons de 5 générations de communes traitées en 2016, 2017, 2018, 2019 et 2020. Nous appelons "cohorte x " l'ensemble de communes ayant reçu le traitement la même année x . La cohorte 2020 n'est jamais traitée dans notre échantillon d'estimation puisque ce chapitre se concentre sur les années 2011-2019.

La cohorte 2020 fait donc partie, dans le cadre de cette analyse, du groupe de contrôle. Se pose alors la question de quelles autres communes inclure dans le groupe de contrôle.

On peut distinguer deux groupes additionnels : (i) les communes ayant déjà reçu le traitement en 2015 et (ii) celles qui le recevront après 2020. En présence d'effets dynamiques, l'inclusion du premier groupe est potentiellement problématique car elle risque de biaiser l'estimateur de double-différence (Goodman-Bacon, 2021). Nous décidons donc d'exclure ce groupe. L'inclusion du deuxième groupe ne pose pas, a priori, de problème de biais lié à des effets dynamiques. Il est possible cependant que ce groupe regroupe des communes très hétérogènes. Ce n'est pas un problème en soit, la question clé étant de savoir si l'hypothèse que ce groupe aurait suivi en moyenne une tendance parallèle à celles des villes traitées est plausible ou non.

Afin d'examiner la validité de cette hypothèse on peut simplement voir si les tendances pré-traitement semblent comparable. Nous présentons des résultats à cet effet ci-dessous.

Nous estimons l'équation (1.3) en considérant deux variable d'intérêts impor-

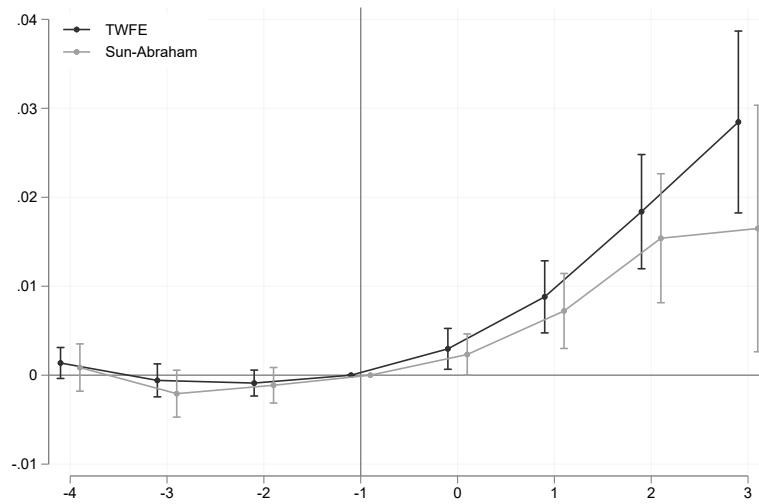
tantes : le stock d'établissement et l'emploi salarié.

La figure 1.13 présente les résultats pour le stock d'établissement. Le panel (a) présente les estimations sur les cohortes 2016 à 2020. La série en noir correspond aux estimations par la biais d'une estimation MCO avec deux ensembles d'effets fixes (année et commune). La série en gris correspond aux estimations utilisant l'estimateur développé par Sun et Abraham (2021) qui permet de prendre en compte l'hétérogénéité des effets de traitements entre cohortes. On constate que les pré-tendances sont relativement faibles et que les deux estimateurs aboutissent à des résultats très proches, suggérant un effet autour de +2 % sur le stock d'établissement trois ans après le déploiement. Le panel (b) montre des résultats issus d'estimations identiques à l'exception de la définition de l'échantillon d'estimation qui inclut désormais également l'ensemble des communes non traitées en 2020 comme groupe de contrôles. Si l'addition de ces observations permet des gains de précision, on voit néanmoins que les tendances pré-traitement sont très différentes. Cette divergence met en doute la capacité des estimations à isoler un effet causal du déploiement de la fibre. Le même constant émerge dans le cas de l'emploi salarié présenté dans la figure 1.14.

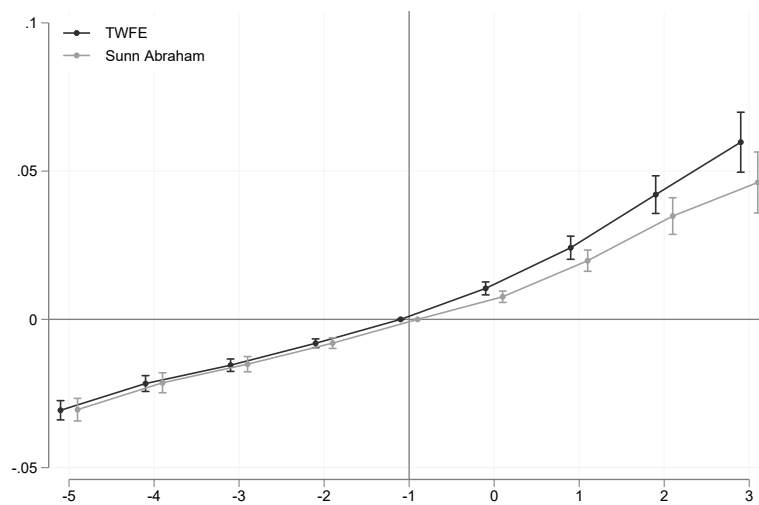
Dès lors, sur la base de ce constat l'analyse empirique de ce chapitre se fondera principalement sur l'échantillon restreint aux cohortes 2016 à 2020.

FIGURE 1.13 – Stock d'établissements

(a) Groupe de contrôle incluant seulement les communes traitées en 2020.



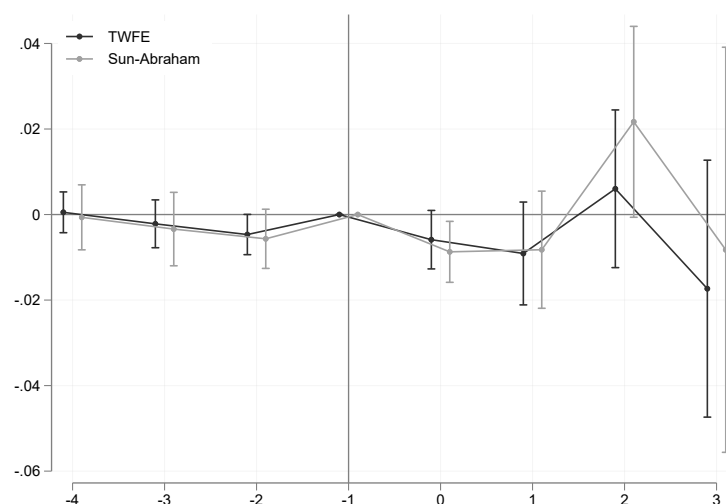
(b) Groupe de contrôle incluant les communes encore non-traitées en 2020.



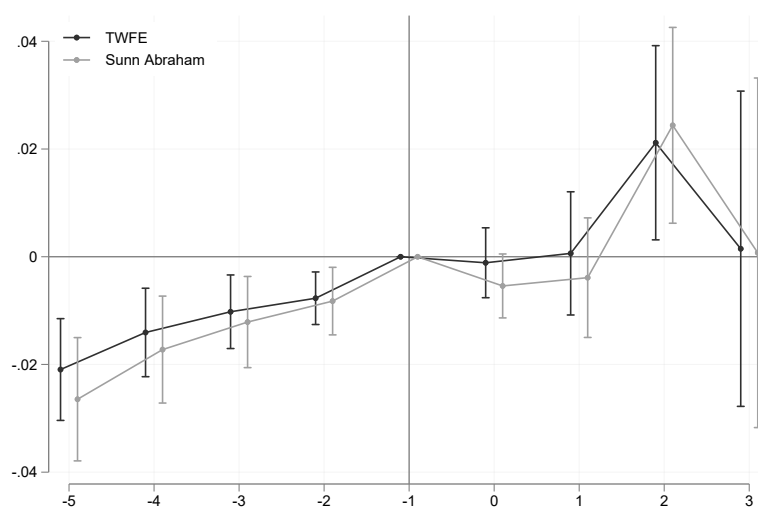
Notes : Cette figure présente les coefficients d'une étude d'évènements ou différence-de-différences dynamiques (voir équation 1.3). Les intervalles de confiance à 95% sont représentés sur la base d'écart-types robustes clustérisés au niveau commune. La variable dépendante est le log du stock d'établissement (source REE). Les coefficients sont normalisés à $t-1$ et $t-5$ pour le panel (a) et $t-1$ pour le panel (b).

FIGURE 1.14 – Emploi salarié

(a) Groupe de contrôle incluant seulement les communes traitées en 2020.



(b) Groupe de contrôle incluant les communes encore non-traitées en 2020.



Notes : Cette figure présente les coefficients d'une étude d'évènements ou différence-de-différences dynamiques (voir équation 1.3). Les intervalles de confiance à 95% sont représentés sur la base d'écart-types robustes clustérisés au niveau commune. La variable dépendante est le log de salariés au 31 décembre (source Acoess). Les coefficients sont normalisés à $t - 1$ et $t - 5$ pour le panel (a) et $t - 1$ pour le panel (b).

CHAPITRE 2

ACCESSIBILITÉ DE LA FIBRE ET RÉSILIENCE À LA CRISE SANITAIRE

2.1 Introduction

Les réseaux à très haut débit sont considérés comme un facteur clé du développement socio-économique. Leur déploiement a été rendu prioritaire dans le monde entier et est considéré comme un investissement pour l'avenir¹.

Plusieurs arguments portent à croire que l'utilité de ce type de réseau a été démultipliée lors de la crise sanitaire liée à la pandémie de Covid-19. Dans un contexte de restrictions sanitaires fortes qui entravent le fonctionnement normal de l'économie en limitant les interactions en face-à-face, de nombreuses activités, habituellement réalisées "en présentiel" (c'est-à-dire impliquant une présence sur un lieu de travail hors domicile, ou une interaction physique interpersonnelle), ont basculé en ligne. Un grand nombre de salariés n'ont ainsi pas pu se rendre sur leur lieu de travail habituel et ont été invités à travailler depuis leur domicile. Pour une grande partie de ces salariés, cette adaptation nécessite une usage accru des

1. voir par exemple la promotion par l'Union Européenne d'une société européenne du gigabit; <https://digital-strategy.ec.europa.eu/en/library/connectivity-european-gigabit-society-brochure>

technologies numériques. Dès lors, la qualité de connexion dont disposent à la fois les entreprises et les salariés dans leur domicile est cruciale afin de permettre qu'une telle adaptation se fasse avec le moins de conséquences néfastes possibles sur les salariés et l'activité économique de l'entreprise.

L'objet de ce chapitre est de documenter les interactions entre la disponibilité de la fibre au niveau local et les effets de la crise sanitaires sur le marché du travail et le marché immobilier. En particulier, la question de recherche est donc de savoir si le très haut débit (THD) a permis d'amortir une partie des effets négatifs de la crise.

A cette fin, nous comparons les dynamiques des marchés du travail et immobilier entre 2019 et 2020, entre des villes couvertes et non-couvertes par la fibre.

Nous constatons plusieurs associations positives entre la dynamique du marché du travail et la couverture fibre début 2020 (juste avant que n'éclate la crise). Ces villes voient une baisse moins prononcée du nombre d'offres d'emploi au 2ème trimestre 2020 postées par des établissements localisés en leur sein. Nous confirmons que cette dynamique se traduit également par une baisse plus faible du nombre d'embauches effectives.

Afin de soutenir l'aspect causal des différences qui émergent entre ces deux groupes de communes, nous nous concentrons sur un sous-ensemble de communes qui, un an avant la crise (au premier trimestre 2019), ne disposaient pas encore de la fibre. Dès lors, on n'identifie comme traitées que les communes couvertes par la fibre juste avant la crise sanitaire et qui ne le sont que du fait du déploiement récent de la fibre et non pas en raison de facteurs structurels (appartenance à une zone très dense, fort dynamisme économique). Nous montrons de plus que nos résultats sont très peu sensibles à l'inclusion d'un nombre important de contrôles observables qui prédisent fortement l'ampleur locale du choc Covid, notamment la part des secteurs exposés (restauration, hôtellerie, animation) dans l'emploi salarié en 2019. Les différences entre communes couvertes et non-couvertes persistent également

lorsque l'on contrôle de manière flexible, via un ensemble d'effets fixes, pour le caractère rural, de montagne ou encore la densité des communes. Enfin, nous montrons que les communes couvertes et non couvertes, suivaient des tendances très comparables avant la crise, entre 2019 et 2018, ce qui justifie l'hypothèse qu'elles auraient suivies des tendances comparables en 2020 en l'absence du déploiement supplémentaire de la fibre entre 2019 et 2020.

Sur la base de données d'enquête individuelles, nous examinons si le télétravail est devenu plus fréquent lors de la crise dans les communes couvertes. Nous adoptons à cette fin une stratégie d'identification différente en exploitant le fait que certains secteurs se prêtent plus ou moins au télétravail, du fait de leur intensité en tâches numériques. Dans cette approche au niveau individuel, nous nous concentrons sur l'effet d'interaction entre la disponibilité de la fibre et la télé-travaillabilité des secteurs d'emploi. Nous trouvons précisément un effet d'interaction positif.

Ce chapitre est structuré de la façon suivante. La section 2.2 présente les bases de données. La section 2.3 présente la stratégie empirique ainsi que la spécification économétrique. Nous présentons des statistiques descriptives en section 2.4 avant de présenter les résultats principaux en section 2.5. La section 2.6 présente les résultats complémentaires sur la base de données individuelles.

2.2 Données

Données d'offres d'emploi. Nous mobilisons des données sur les offres d'emploi provenant de pole-emploi.fr, le site de publication d'offres d'emploi géré par le service public de l'emploi français (Pôle Emploi). Toute entreprise privée peut utiliser ce site pour publier des annonces d'emploi et filtrer les profils des demandeurs d'emploi gratuitement. Jusqu'en 2021, pour chaque offre d'emploi, nous disposons d'informations sur l'identifiant de l'établissement qui la publie, l'intitulé du poste, la profession, les horaires de travail, le lieu de travail, la date de publication et le

saire annoncé². Pôle Emploi est l'un des principaux sites de publication d'offres d'emploi en France, couvrant près de 50 % des embauches réalisées par le biais d'annonces en ligne—d'après une enquête de 2016 au niveau des entreprises sur les pratiques de recrutement (OFER).

Nous utilisons jusqu'à présent seulement la localisation et la date de l'offre d'emploi.

Données d'embauches : DPAE. Tous les employeurs sont tenus de déclarer toute embauche d'un salarié relevant du régime général de la Sécurité sociale dans les huit jours précédant le début du contrat. Pour cela, ils doivent adresser une "Déclaration préalable à l'embauche" (DPAE) à l'Urssaf dont dépend l'établissement concerné. Les données de la source DPAE sont issues de ces déclarations. Elles contiennent les informations relatives aux établissements (SIRET, nom ou raison sociale, adresse et APE), aux salariés embauchés (sexe, date de naissance) et les principales caractéristiques du contrat de travail (date et heure d'embauche, type de contrat, date de fin de CDD le cas échéant).

Nous apparions les DPAE au répertoire Sirène afin de déduire la localisation (code commune Insee) de l'établissement.

Immobilier. La Direction Générale des Finances Publiques (DGFIP) a développé et rendu public le fichier "Demande de Valeurs Foncières" (DVF) qui recense l'ensemble des mutations foncières à titre onéreux publiées dans les services de la publicité foncière. Cette base de données fiscales est structurée, vérifiée et enrichie par le Cerema qui produit la base de données DV3F, qui contient des données foncières complémentaires issues des fichiers fonciers.

Nous utilisons ici cette base afin de calculer le nombre de mutations par com-

2. Ce dernier est complété par un texte qui peut apporter des informations supplémentaires sur la fixation du salaire et la rémunération (par exemple, si la rémunération est liée à l'expérience ou à la qualification, ou si elle fait l'objet d'une négociation).

mune et année.

Emploi. La structure de l'emploi au niveau communal est issues des fichiers de l'Acoss. L'Acoss met en ligne les dénombrements des établissements employeurs et des effectifs salariés, par catégories d'activités (nomenclature NAF à 732 postes) et par commune, au 31 décembre. Nous utilisons ces données afin d'évaluer l'évolution de l'emploi ainsi que pour construire des variables de contrôles mesurant la part dans les économies locales des secteurs très exposés à la crise. Ces secteurs sont les suivants : Commerce de détail, Hébergement, Restauration, Activités liées à l'emploi (inclut l'intérim).

Recensement. De la même manière, nous complétons la description de la structure des économies locales par plusieurs fichiers issus du recensement. Tout d'abord, les caractéristiques socio-démographiques (CSP, part de retraités) en 2017 proviennent du recensement de l'Insee. D'autres caractéristiques des communes ont également été extraites des bases de données de l'Insee, comme la catégorie de densité de population, le nombre de ménages, le nombre de logements, le degré d'urbanisation de la commune et son caractère montagnoux. Ces informations sont évaluées en 2017.

2.3 Stratégie empirique

La question de recherche qui motive ce chapitre est la suivante : le déploiement du THD a-t-il permis d'amortir une partie des effets négatifs de la crise ? Pour répondre empiriquement à cette question, nous estimons une équation en premières différences qui vise à tester si l'effet du choc Covid est différencié entre villes très couvertes et peu couvertes par la fibre au premier trimestre 2020 (2020T1).

L'équation d'estimation est la suivante :

$$\Delta Y_i = \alpha + \delta \times \text{Fibre}_i + X_i' \beta + \psi_{d(i)} + \epsilon_i, \quad (2.1)$$

où : ΔY_i est l'évolution de la variable d'intérêt pour la commune i entre 2019 et 2020 (pour un trimestre donné), X_i sont des caractéristiques de la commune, $\psi_{d(i)}$ un ensemble d'effets fixes (rural, montagne, densité). La variable Fibre_i est la variable explicative d'intérêt. Il s'agit d'une indicatrice = 1 si le taux de couverture fibre $\geq 75\%$, et 0 sinon.

2.4 Statistiques descriptives

Dans cette section nous présentons des statistiques descriptives sur les communes incluses dans notre échantillon d'estimation. Le tableau 2.1 présente ces statistiques en se focalisant sur les communes dont le taux de couverture fibre était nul au début de 2019. La colonne (1) montre la moyenne pour les communes dont la couverture est faible au premier trimestre 2020 ($< 75\%$ des locaux), la colonne (2) montre la moyenne des communes avec une forte couverture à la même date ($\geq 75\%$). La colonne (3) présente la différences entre les deux groupes. On constate que le groupe de contrôle (colonne 1) contient un nombre beaucoup plus important de communes (environ 16 500) que le groupe de traitement (environ 1 000). Les deux groupes présentent par définition des taux de couverture très différents, bien que le groupe de contrôle n'ait pas un taux de couverture nul.

Les deux groupes contiennent principalement des communes relevant de RIP (réseau d'initiative publique) et ces communes sont légèrement plus représentées dans le groupe de traitement.

TABLEAU 2.1 – Statistique descriptives selon couverture au 2020T1 : échantillon non couvert en 2019T1

	Fibre = 0	Fibre = 1	Différence
<i>Couverture</i>			
Part couverture fibre	0,0399	0,9119	0,8720***
Commune zone RIP	0,8234	0,8646	0,0411***
<i>Structure population</i>			
Part de cadres	0,0691	0,0835	0,0144***
Part d'agriculteurs	0,0389	0,0278	-0,0110***
Secteurs forte exposition crise	0,1610	0,1426	-0,0184**
Population 2017	1339	1111	-227***
Log Population 2017	6,5844	6,4528	-0,1315***
<i>Type de commune</i>			
Catégorie de densité (1 à 4)	3,1207	3,0298	-0,0909***
Commune rurale	0,7679	0,7618	-0,0062
Commune de montagne	0,1820	0,0788	-0,1032***
<i>Tendances 2018-2019</i>			
Moyenne $\Delta_{t,t-4} \ln$ (offres) 2019-2018	0,0338	0,0121	-0,0217
N	16654	1041	

Notes : Ce tableau présente la moyenne de certaines variables mesurées entre 2017 et 2019 selon le degré de couverture fibre de chaque ville au premier trimestre de 2020. La variable binaire "Fibre" qui définit le groupe de traitement est égale à 1 lorsque la part de locaux d'une ville disposant d'un accès à la fibre est supérieur à 75%. La variable est sinon égale à 0. La variable de densité est une variable ordonnée de 1 à 4 (1 : densément peuplée 2 : densité intermédiaire 3 : peu dense 4 : très peu dense).

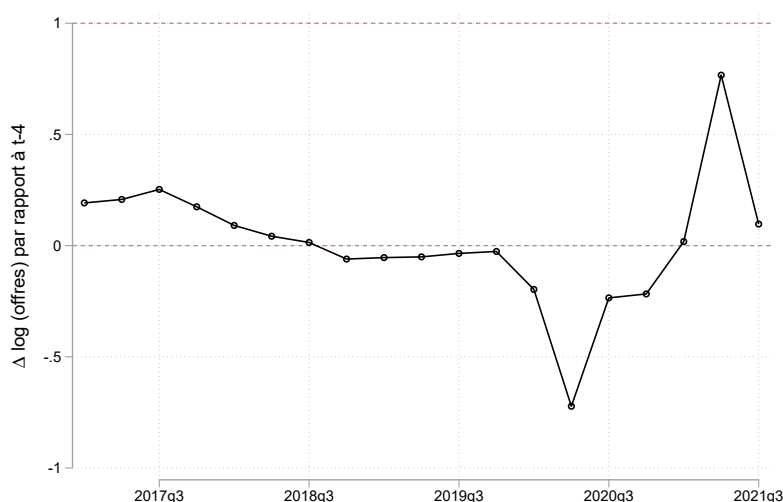
La part des communes rurales est très importante dans les deux groupes, et n'est pas significativement différente. La part des secteurs exposés à la crise est légèrement plus forte au sein des communes non-couvertes. Cette différence est statistiquement significative bien qu'elle ne soit pas quantitativement très importante (1,8 pp alors que la moyenne dans le groupe traité est de 14,3 %). Au vu de ces différences de structure, nous introduirons ces caractéristiques comme variables de contrôles dans les estimations, afin de contrôler leur effet potentiel sur les écarts de dynamiques estimés sur le marché du travail et de l'immobilier.

Cette restriction aux communes ayant une couverture fibre nulle au début de 2019 implique une focalisation sur des communes plus rurales et plus susceptibles

de relever de l'initiative publique³, et rend ainsi notre analyse plus pertinente vis-à-vis des déploiements futurs de la fibre.

La figure 2.1 présente le taux de croissance entre les trimestres t et $t - 4$ du nombre d'offres postées. On constate une croissance légèrement positive puis très proche de zéro jusqu'au 1er trimestre de 2020, où le nombre d'offres se contracte d'environ 10 % avant de s'effondrer avec un taux de croissance de -70 % au deuxième trimestre de 2020. On constate un léger rétablissement aux troisième et quatrième trimestres, sans que les taux de croissance ne redeviennent positifs en 2020. Les taux de croissance sur 2021 sont très positifs, ce qui reflète en partie la faible valeur du dénominateur (le nombre d'offres postées au cours des différents trimestres de 2020).

FIGURE 2.1 – Dynamique des offres d'emplois : $\Delta_{t,t-4} \ln(\text{offres})$



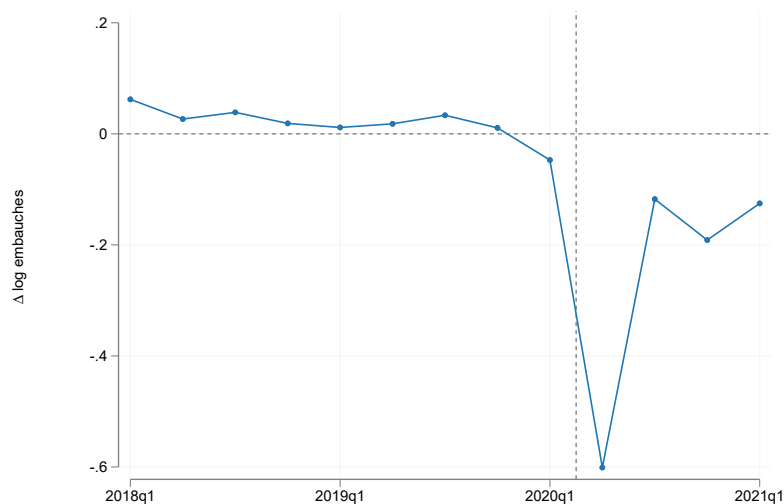
Notes : Cette figure montre le taux de croissance annuel par trimestre du nombre d'offres d'emplois.

La figure 2.2 présente le taux de croissance entre les trimestres t et $t - 4$ du nombre d'embauches effectives telles que documentées dans les DPAE. On constate

3. En effet, en n'appliquant pas cette restriction, les différences de caractéristiques observables entre communes couvertes et non-couvertes par la fibre au premier trimestre 2020 sont beaucoup plus prononcées (voir en annexe le tableau 2.9, qui reproduit les comparaisons du tableau précédent, toutes communes confondues).

une forte contraction au deuxième trimestre de 2020 d'un ampleur comparable à la baisse du nombre d'offres. Les taux de croissance demeurent négatifs pour 2020 et 2021T1.

FIGURE 2.2 – Dynamique des offres d'emplois : $\Delta_{t,t-4} \ln(\text{embauches})$

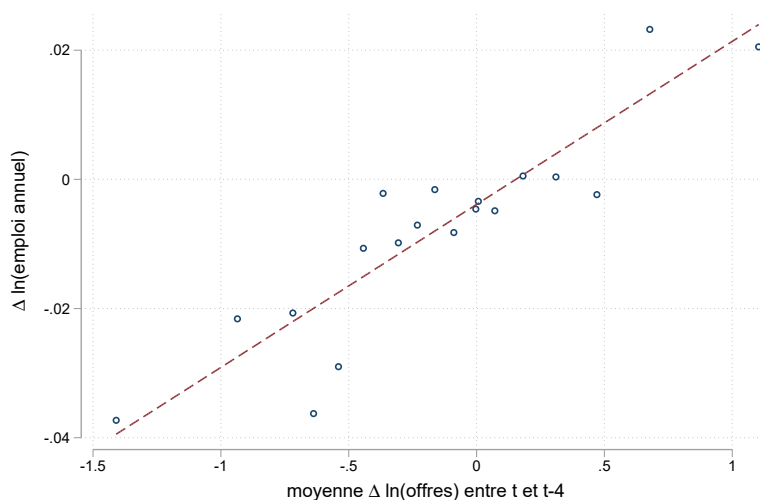


Notes : Cette figure montre le taux de croissance annuel par trimestre du nombre d'embauches (source DPAE).

Globalement, ces deux figures montrent que les bases de données mobilisées captent bien l'ampleur et la dynamique de la crise Covid sur le marché du travail.

La figure 2.3 montre la relation entre la croissance moyenne du nombre d'offres postées chaque trimestre en 2020 et l'évolution du nombre d'emplois à la fin de l'année. On constate une relation positive nette.

FIGURE 2.3 – Corrélation entre croissance moyenne du nombre d'offres postées chaque trimestre en 2020 ($\Delta_{t,t-4}$) et croissance de l'emploi au 31/12 ($\ln(\text{offres})$ et $\Delta_{19,20} \ln(\text{emploi})$)



Notes : Cette figure montre la relation entre, d'une part, la moyenne en 2020 du taux de croissance annuel par trimestre du nombre d'offres d'emplois (source Pôle emploi) et, d'autre part, la croissance de l'emploi au 31/12 (source Acoess).

2.5 Résultats principaux

Nous présentons ici les principaux résultats de notre analyse sur le marché du travail pendant la crise.

2.5.1 Dynamique des offres d'emploi et des embauches (2020T2)

Le tableau 2.2 montre l'impact différentiel de la crise sur la dynamique des offres d'emploi en nous concentrant sur le taux de croissance entre 2020T2 et 2019T2, c'est-à-dire sur le trimestre le plus affecté par les restrictions d'activité liées au contexte sanitaire.

La colonne (1) montre qu'une forte couverture fibre est associée à un taux de croissance de +10,7 pp. Colonne (2) procède à l'intégration de trois variables de contrôles continues : la part de cadres, la part d'agriculteurs et la part des sec-

teurs à fortes exposition à la crise. Ces variables sont des prédicteurs potentiels de la sévérité des effets économiques de la crise sanitaire dont on sait qu'ils ont été forts dans certains secteurs et ainsi que dans les zones denses urbaines par rapport aux zones rurales et périurbaines. Ces variables sont des prédicteurs statistiquement significatifs de la sévérité de la crise. C'est en particulier le cas de la part des secteurs exposés dont le coefficient est très précisément estimé et implique qu'une hausse de 10 pp de la part de de ces secteurs prédit une baisse de 3,3 pp ($0,10 \times (1 - \exp(-0,429)) = -0,033$) du taux de croissance des offres d'emploi. Les communes avec une forte part d'agriculteurs et une faible part de cadres ont eu tendance à connaître des taux des taux de croissance plus élevés. Ces deux coefficients, auxquels nous ne donnons pas une interprétation causale, reflètent vraisemblablement le fait que la crise sanitaire ait dans un premier temps affecté les centres urbains plus fortement que les communes rurales ou périurbaines qui sont moins denses.

En dépit du pouvoir prédictif de ces trois variables, on note que le coefficient associé de couverture est très stable et légèrement plus élevé (+12,3 pp). C'est également le cas lorsque l'on introduit, dans colonne (3), des effets fixes pour la catégorie de densité et le caractère rural ou bien montagnard des communes (+11,4 pp). Enfin, la colonne (4) présente le résultat d'une estimation similaire où la variable dépendante est retardée d'un an, c'est à dire que l'on examine, dans l'esprit d'un test placebo, l'effet du déploiement de la fibre en 2020 sur la trajectoire de croissance des offres d'emploi entre 2019T2 et 2018T2. Le coefficient associé à la couverture fibre au premier trimestre 2020 devient très proche de 0 et est non significativement différent de 0, ce qui valide par contraste notre stratégie sur l'évolution des offres d'emploi entre 2019T2 et 2020T2.

TABLEAU 2.2 – Évolution des offres d'emploi 2020T2

	$\Delta_{t,t-4} \ln(\text{offres})$ T2 2019 2020			— 2019-2018
	(1)	(2)	(3)	(4)
Forte Couverture Fibre 2020T1	0,107*** (0,0407)	0,123*** (0,0402)	0,114*** (0,0403)	-0,0107 (0,0367)
Part de cadres		-0,479*** (0,177)	-0,239 (0,181)	0,0253 (0,169)
Part d'agriculteurs		1,603*** (0,260)	0,550* (0,293)	-0,647*** (0,239)
Secteurs forte exposition crise		-0,429*** (0,0496)	-0,396*** (0,0500)	0,0730* (0,0404)
R^2	0,001	0,018	0,027	0,003
Adjusted R^2	0,001	0,017	0,026	0,002
Effet fixe densité + rural + mont, Placebo			✓	✓
Observations	9303	9158	9158	10349

Notes : La variable dépendante est le taux de croissance du nombre d'offres d'emploi entre 2020T2 et 2019T2. La colonne (4) présente un test de pré-tendance commune en utilisant le taux de croissance entre 2019T2 et 2018T2 comme variable dépendante. L'échantillon d'estimation inclut l'ensemble des communes n'ayant aucune couverture fibre en 2019T1. Les communes définies comme ayant une forte couverture au moment de la crise Covid sont celles ayant une couverture supérieure à 75% au premier trimestre de 2020 selon les données de l'Arcep sur le déploiement de la fibre. Les variables de contrôle sont les suivantes : *Secteurs forte exposition crise* calcule la part de l'emploi dans des secteurs fortement affectés par la crise en 2019. Parmi les effets fixes : *densité* fait référence aux catégories de densité définies par l'Insee (4 valeurs), rural et montagne sont des variables binaires. Les contrôles démographiques (CSP) contrôlent pour la part de cadre et de retraités dans la population en 2017-2018. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

On constate ainsi que les communes avec une forte couverture fibre ont significativement moins souffert de la crise : les établissements situés dans ces communes ont moins diminué les offres postées. La robustesse de cet écart à l'introduction de contrôles pertinents ainsi que l'existence d'une tendance pré-crise commune aux localités couvertes et non-couvertes supporte l'interprétation causale de ce coefficient. Le taux de croissance moyen au sein de l'échantillon d'estimation étant d'environ -0,40, la couverture fibre permet donc de réduire l'ampleur du choc d'environ 25 % par rapport à sa valeur moyenne (10/40).

Au-delà des offres postées, il est intéressant de déterminer si les embauches effectives sont également affectées. Le tableau 2.3 montre que c'est effectivement le cas. On voit dans la colonne (3) que la couverture fibre est associée (après avoir contrôlé pour l'ensemble des variables observables) à un taux de croissance entre

2020T2 et 2019T2 de 8,3 pp supérieur. Compte tenu de la baisse moyenne de -0,587, cela implique que la fibre permet d’amortir le choc à hauteur de 14 % de sa valeur moyenne.

TABLEAU 2.3 – Évolution des embauches 2020T2

	$\Delta \ln(\text{embauches})_{t,t-4}$		T2 2019-20	— 2019-2018
	(1)	(2)	(3)	(4)
Forte Couverture Fibre 2020T1	0.0670** (0.0330)	0.0746** (0.0326)	0.0826** (0.0327)	0.0152 (0.0383)
Part de cadres		-0.730*** (0.151)	-0.699*** (0.154)	0.0625 (0.173)
Part d’agriculteurs		-0.0294 (0.203)	-0.200 (0.233)	-0.406* (0.235)
Secteurs forte exposition crise		-0.454*** (0.0450)	-0.488*** (0.0460)	0.113*** (0.0417)
R^2	0.000	0.013	0.016	0.001
Adjusted R^2	0.000	0.013	0.015	0.000
Effet fixe densité + rural + mont.			✓	✓
Placebo				✓
Observations	13852	13397	13397	15126

Notes : Cette table présente les résultats d’une régression en première différence (voir équation (3.1) pour plus de détails). La variable dépendante est le taux de croissance du nombre d’embauches entre 2020T2 et 2019T2. La colonne (4) présente un test de pré-tendance commune entre utilisant le taux de croissance entre 2019T2 et 2018T2 comme variable dépendante. L’échantillon d’estimation inclut l’ensemble des communes n’ayant aucune couverture fibre en 2019T1. Les communes définies comme ayant une forte couverture au moment de la crise Covid sont celles ayant une couverture supérieure à 75% au premier trimestre de 2020 selon les données de l’Arcep sur le déploiement de la fibre. Les variables de contrôles sont les suivantes : *Secteurs forte exposition crise* calcule la part de l’emploi dans des secteurs fortement affectés par la crise en 2019. Parmi les effets fixes : *densité* fait référence aux catégories de densité définies par l’Insee (4 valeurs), rural et montagne sont des variables binaires. Les contrôles démographiques (CSP) contrôlent pour la part de cadre et de retraités dans la population en 2017-2018. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Le tableau 2.4 présente les résultats pour des taux de croissance calculés pour différents trimestres. La spécification retenue correspond à la colonne (3) des tableaux 2.2 et 2.3 (avec effets fixes et contrôles observables). Le tableau présente les résultats pour les offres et les embauches. On constate globalement que l’effet est concentré au 2^{ème} semestre de 2020 pour les deux variables (colonne 2). On constate dans le cas des embauches néanmoins un léger déclin au 2020T3 ainsi qu’un effet de nouveau positif en 2021T1.

Notons que le deuxième semestre de 2021 pose des problèmes d’interprétation dans la mesure où il s’agit de la première période pour laquelle la période de ré-

férence utilisée afin de calculer le taux de croissance est déjà affectée par la crise. Le fait que la part de secteurs à forte exposition prédise *positivement* la croissance à la fois des offres et des embauches illustre l'aspect atypique de cette période par rapport aux autres, mais relève avant tout d'un artefact statistique (les résultats de la colonne (2) du tableau soulignent bien que les dénominateurs du taux de croissance, examiné dans la colonne (6), sont différentiellement affectés par la crise selon que les communes sont couvertes ou non-couvertes par la fibre).

TABLEAU 2.4 – Évolution des offres et embauches 2020T1 à 2021T1

	(1) 2020q1	(2) 2020q2	(3) 2020q3	(4) 2020q4	(5) 2021q1	(6) 2021q2
Offres						
Forte Couverture Fibre 2020T1	0,0305 (0,0374)	0,114*** (0,0403)	0,0574 (0,0359)	0,0756* (0,0403)	-0,00750 (0,0369)	-0,0751* (0,0398)
Secteurs forte exposition crise	-0,115*** (0,0425)	-0,396*** (0,0500)	0,0113 (0,0438)	-0,457*** (0,0562)	-0,420*** (0,0454)	0,486*** (0,0493)
Adjusted R^2	0,002	0,026	0,004	0,017	0,012	0,027
Moyenne dép, var	-0,08	-0,40	-0,08	-0,10	0,07	0,60
Observations	9716	9158	9959	8595	9537	9380
Embauches						
Forte Couverture Fibre 2020T1	0,00909 (0,0267)	0,0826** (0,0327)	-0,0560** (0,0259)	0,00891 (0,0301)	0,0766** (0,0327)	-0,0282 (0,0302)
Secteurs forte exposition crise	-0,128*** (0,0333)	-0,488*** (0,0460)	-0,0615* (0,0325)	-0,665*** (0,0430)	-0,738*** (0,0464)	0,231*** (0,0360)
Adjusted R^2	0,001	0,015	0,000	0,027	0,032	0,005
Moyenne dép, var,	-0,048	-0,587	-0,102	-0,190	-0,141	0,486
Observations	13692	13397	14616	13477	13603	13463
Effet fixe densité + rural + mont,	✓	✓	✓	✓	✓	✓
Contrôles CSP	✓	✓	✓	✓	✓	✓

Notes : Cette table présente les résultats d'une régression en première différence (voir équation (3.1) pour plus de détails). La variable dépendante est le taux de croissance annuel du nombre d'offres ou d'embauches pour un trimestre donné (2020T1 à 2021T2) et un an auparavant. L'échantillon d'estimation inclut l'ensemble des communes n'ayant aucune couverture fibre en 2019T1. Les communes définies comme ayant une forte couverture au moment de la crise Covid sont celles ayant une couverture supérieure à 75% au premier trimestre de 2020 selon les données de l'Arcep sur le déploiement de la fibre. Les variables de contrôles sont les suivantes : *Secteurs forte exposition crise* calcule la part de l'emploi dans des secteurs fortement affectés par la crise en 2019. Parmi les effets fixes : *densité* fait référence aux catégories de densité définies par l'Insee (4 valeurs), *rural* et *montagne* sont des variables binaires. Les contrôles démographiques (CSP) contrôlent pour la part de cadre et de retraités dans la population en 2017-2018. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Le tableau 2.5 montre le même type de résultats, cette fois sur la croissance de l'emploi annuel au 31 décembre. Si les différences sont positives quelque soient les spécifications de (1) à (3), elles sont néanmoins d'une ampleur économique beau-

coup plus faible et estimées peu précisément. On ne détecte ainsi pas d'effet précis sur l'emploi de la couverture fibre. Ce résultat à première vue surprenant s'explique par le recours aux dispositifs d'activité partielle : les salariés en activité partielle sont considérés comme en emploi, et compte tenu du contexte de recours massif à l'activité partielle, l'emploi n'a été que faiblement affecté eu égard à l'ampleur du choc macroéconomique. Ainsi, cette absence d'effet sur le stock d'emploi souligne l'intérêt de s'intéresser aux flux d'offres et d'embauches afin de capter la dynamique du marché du travail pendant la crise Covid et comment cette dynamique a interagi avec la disponibilité de la fibre (dans un contexte de recours massif au télétravail).

TABLEAU 2.5 – Évolution d'emploi au 31/12 entre 2020 et 2019 :
 $\Delta_{19,20} \ln(\text{emploi})$

	moyenne $\Delta_{2019,2020} \ln(\text{emploi})$ 2019 2020			— 2019-2018
	(1)	(2)	(3)	(4)
Forte Couverture Fibre 2020T1	0,00791 (0,00747)	0,00520 (0,00738)	0,00298 (0,00742)	0,0110 (0,0103)
Part de cadres		0,0607 (0,0375)	0,0363 (0,0390)	0,0828** (0,0384)
Part d'agriculteurs		0,0543 (0,0546)	0,108* (0,0625)	-0,0530 (0,0667)
Secteurs forte exposition crise		-0,127*** (0,0141)	-0,123*** (0,0141)	0,00971 (0,0123)
R^2	0,000	0,012	0,014	0,001
Adjusted R^2	-0,000	0,012	0,014	0,000
Effet fixe densité + rural + mont, Placebo			✓	✓
Observations	15030	15030	15030	15518

Notes : Cette table présente les résultats d'une régression en première différence (voir équation (3.1) pour plus de détails). La variable dépendante est le taux de croissance annuel du nombre d'offres ou d'embauches pour un trimestre donné (2020T1 à 2021T2) et un an auparavant. L'échantillon d'estimation inclut l'ensemble des communes n'ayant aucune couverture fibre en 2019T1. Les communes définies comme ayant une forte couverture au moment de la crise Covid sont celles ayant une couverture supérieure à 75% au premier trimestre de 2020 selon les données de l'Arcep sur le déploiement de la fibre. Les variables de contrôles sont les suivantes : *Secteurs forte exposition crise* calcule la part de l'emploi dans des secteurs fortement affectés par la crise en 2019. Parmi les effets fixes : *densité* fait référence aux catégories de densité définies par l'Insee (4 valeurs), *rural* et *montagne* sont des variables binaires. Les contrôles démographiques (CSP) contrôlent pour la part de cadre et de retraités dans la population en 2017-2018. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.5.2 Marché immobilier

Le tableau 2.6 présente les résultats d'estimations relatives au marché de l'immobilier, considérant comme variable dépendante le taux de croissance du nombre de mutations. Le principal enseignement des ces résultats est un effet positif de la couverture par la fibre, ici encore stable selon les spécifications choisies et robuste à l'introduction de variables de contrôles, et ici encore confirmé par l'absence de tendances différentielles sur la période 2018-2019. L'effet d'un passage à un taux de 100% de couverture fibre est associé à une augmentation d'environ 18% du nombre de mutations sur la période.

TABLEAU 2.6 – Évolution du nombre de mutations (transactions immobilières) 2019-2020

	$\Delta \ln(\text{mutations})$ 2019 2020			— 2019-2018
	(1)	(2)	(3)	(4)
Forte Couverture Fibre 2020T1	0,178*** (0,0417)	0,185*** (0,0421)	0,165*** (0,0418)	-0,0280 (0,0333)
Part de cadres		0,488*** (0,187)	0,647*** (0,189)	-0,498*** (0,143)
Part d'agriculteurs		2,453*** (0,317)	1,251*** (0,363)	0,00353 (0,236)
Secteurs forte exposition crise		-0,272*** (0,0563)	-0,212*** (0,0562)	-0,0515 (0,0386)
R^2	0,002	0,012	0,023	0,009
Adjusted R^2	0,002	0,012	0,022	0,009
Effet fixe densité + rural + mont, Placebo			✓	✓
Observations	10202	10082	10082	13800

Notes : La variable dépendante est le taux de croissance du nombre de mutations au niveau communal entre 2020 et 2019. La colonne (4) présente un test de pré-tendance commune entre utilisant le taux de croissance entre 2019 et 2018 comme variable dépendante. L'échantillon d'estimation inclut l'ensemble des communes n'ayant aucune couverture fibre en 2019T1. Les communes définies comme ayant une forte couverture au moment de la crise Covid sont celles ayant une couverture supérieure à 75% au premier trimestre de 2020 selon les données de l'Arcep sur le déploiement de la fibre. Les variables de contrôles sont les suivantes : *Secteurs forte exposition crise* calcule la part de l'emploi dans des secteurs fortement affectés par la crise en 2019. Parmi les effets fixes : *densité* fait référence aux catégories de densité définies par l'Insee (4 valeurs), rural et montagne sont des variables binaires. Les contrôles démographiques (GSP) contrôle pour la part de cadre et de retraités dans la population en 2017-2018. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.5.3 Tests de robustesse

Nous procédons à deux tests de robustesse pour évaluer la nature causale de l'interprétation de nos résultats. Premièrement, nous modifions la forme fonctionnelle utilisée pour calculer le taux de croissance, afin de mieux capter la marge extensive, c'est-à-dire le fait que certaines communes aient un nombre nul d'embauches ou d'offres postées un trimestre donné. Dans un second temps, nous montrons les résultats de régressions dans lesquelles le groupe de contrôle contient uniquement les communes dont le taux de couverture était nul au début de 2019 (comme précédemment) mais dont le taux de couverture a progressé au cours de l'année 2019. Il s'agit de communes qui comme celles du groupe de traitement ont vu les travaux de déploiement de la fibre commencer au cours de l'année 2019 mais pour lesquels les progrès des travaux ont été plus faibles. L'objectif d'une telle procédure est de rendre les groupes de communes traitées et contrôles encore plus comparables que précédemment, et ainsi de rendre l'hypothèse d'identification causale de nos résultats d'autant plus crédible.

Taux de croissance Davis-Haltiwanger (DH). Le taux de croissance en log n'est défini que si la variable est strictement positive dans les deux périodes considérées. (Davis et Haltiwanger, 1992) propose un taux de croissance qui est bien défini si la variable est positive dans l'une des deux périodes considérées. Le taux de croissance DH (Davis-Haltiwanger) de la variable x entre t et $t - 1$ est défini de la façon suivante :

$$DH_t^x = \frac{x_t - x_{t-1}}{\frac{1}{2}(x_t + x_{t-1})} \quad (2.2)$$

Cette variable est égale à +2 pour les communes avec $x_{t-1} = 0$ et $x_t > 0$ et réciproquement +2 dans le cas inverse ($x_{t-1} > 0$ et $x_t = 0$).

Le tableau 2.12 de l'annexe montre que les résultats sont largement comparables avec cette variable.

Groupe de contrôle alternatif. Nous restreignons le groupe de contrôle de communes qui, comme celles du groupe de traitement, ont vu les travaux de déploiement de la fibre commencer au cours de l'année 2019 mais pour lesquels les progrès des travaux ont été plus faibles. Les résultats reposent sur un groupe de contrôle numériquement beaucoup plus limité. La taille de l'échantillon passe de 9 000 à 2 000 pour les offres (voir tableau 2.13 colonne 3) et de 13 400 à 2 800 pour les embauches (voir tableau 2.14 colonne 3).

On remarque néanmoins que les coefficients estimés sont proches de ceux des estimations précédentes ; malgré la taille plus restreinte de l'échantillon, ils restent très précisément estimés, et leur magnitude est même légèrement plus élevée que précédemment.

2.6 Résultats complémentaires : télétravail et activité partielle sur données individuelles

L'approche empirique que nous développons dans cette section consiste à estimer l'effet de la couverture fibre sur le recours au télétravail et à l'activité partielle lors de la crise Covid, au niveau individuel. Nous examinons ainsi l'effet de résider dans une commune dont la couverture fibre est très élevée sur l'emploi et ses caractéristiques au niveau individuel. Pour ce faire, nous mobilisons principalement l'Enquête Emploi en Continu. Du fait du plus faible nombre d'observations liée à l'utilisation de données d'enquête, nous incluons l'ensemble des communes dans l'échantillon d'estimation.

Il existe de manière triviale un biais de variable omise lié à la comparaison d'individus résidant dans des communes très couvertes et peu couvertes. C'est pourquoi nous développons dans cette section une approche alternative en nous intéressant à l'effet d'interaction entre la disponibilité de la fibre et un indice représentant la

télétravaillabilité du poste occupé par l'individu, défini au niveau de sa PCS (professions et catégories socioprofessionnelles). Cet effet d'interaction peut être estimé en incluant des effets fixes au niveau commune, et en incluant un ensemble très riche de variables de contrôle au niveau individuel (âge, sexe, situation familiale, secteur d'activité, etc). Intuitivement, cette stratégie rend compte des écarts de recours au télétravail entre PCS intensives en outils numériques et PCS peu intensives en outils numériques, et compare les différences dans cet écart entre les individus résidant dans des communes couvertes par la fibre et les individus résidant dans des communes non-couvertes.

2.6.1 Données individuelles

Les données individuelles mobilisées proviennent de deux sources.

L'Enquête emploi en continu (EEC) sur les années 2019 à 2020 nous permet d'obtenir les caractéristiques individuelles (âge, sexe, situation familiale, **commune de résidence**, etc). Elle contient une variable sur la commune de résidence, ce qui permet de construire un indice d'accès à la fibre dont la définition est identique à celle adoptée précédemment. On connaît également les caractéristiques de l'emploi (SIREN, secteur d'activité) ainsi que la PCS.

La situation dans l'emploi est documentée, en particulier il existe plusieurs questions en 2019 et 2020 sur le recours à l'activité partielle ainsi que le fait d'exercer son activité professionnelle à domicile, que nous considérerons comme principalement relevant du télétravail (plusieurs éléments viendront valider cette interprétation).

La principale innovation empirique par rapport aux sections précédentes est de comparer les individus les uns aux autres dans deux dimensions : la disponibilité de la fibre dans leur commune de résidence, et un indice représentant l'importance de l'outil numérique dans leur activité professionnelle. Cet indice est construit sur

la base de l'enquête Conditions de Travail 2019, qui documente la part du temps de travail normal passé sur des outils numériques en 2019. Sur la base de ces données, on construit des indices moyens au niveau PCS 4 caractères (457 catégories).

Nous détaillons la construction de l'index télé-travaillabilité dans l'annexe dans la section 2.8.1. Le résultat de ce travail est une variable $part_info_p$ qui capte la part moyenne prédite du temps travail passée sur des outils numériques au sein de chaque PCS p .

2.6.2 Spécification empirique

Une fois construites les données détaillées dans la section précédente, on estime le modèle suivant :

$$Y_{ismcp} = \delta X_{ipmc} + \zeta_s + \alpha_m + \gamma_c + \xi_p + \beta \text{Fibre}_c \times part_info_p + \epsilon_{ipmc} \quad (2.3)$$

où

- Y_{ismcp} est la variable d'intérêt pour l'individu i , interrogé au mois m , salarié du privé dans le secteur s , résidant dans la commune c et occupant une profession p
- X sont des caractéristiques de l'individu (âge, sexe)
- $\zeta_s, \delta_m, \gamma_c, \xi_p$ une séries d'effets fixes (de mois d'enquête, secteur d'activité A88, commune de résidence c , profession p)
- Fibre_c est une variable indicatrice qui prend la valeur 1 quand le taux de couverture de la commune de résidence en réseau supérieur à la fibre est égal à au moins 75%, et 0 sinon au 2020T1.
- $part_info_p$ est la part moyenne prédite du temps travail passée sur des outils numériques au sein de l'occupation p renseigné par l'individu i .

Les écarts-types robustes à l'hétéroscédasticité sont clusterisés / groupés au niveau commune.

2.6.3 Résultats sur télétravail

Le tableau 2.7 montre les résultats pour le télétravail. On voit dans la colonne (1) que la part_info prédit positivement et précisément le recours au télétravail. On constate que le coefficient lié à la couverture est légèrement négatif mais d'une ampleur très limitée. Néanmoins, l'interaction est entre ces deux variables est fortement positive : le recours au télétravail est d'autant plus fortement associé à la fibre que l'on considère des occupations intensives en télétravail, ou bien entre d'autres termes, la probabilité de télétravailler est d'autant plus forte au sein des occupations intensives en numérique que les employés vivent dans des communes couvertes par la fibre.

La colonne (2) introduit des contrôles sectoriels (effets fixes secteur A88), géographique (effet fixe département) et démographique (effet fixes âge et genre). Le terme d'interaction est atténué mais demeure largement positif et significatif. La colonne (3) introduit, en plus des contrôles déjà présents, des effets fixes au niveau des catégories socio-professionnelles à trois caractères⁴ ; dans cette spécification la corrélation entre la part du temps de travail passé sur les outils numériques et la propension au télétravail est estimée sur la variation résiduelle conditionnelle à un niveau très fin de catégorisation socio-professionnelle ; la corrélation entre la couverture fibre et la propension à télétravailler est quant à elle estimée à départements donnés, c'est-à-dire par la comparaison entre communes d'un même département couvertes et non-couvertes par la fibre⁵.

De la même manière que précédemment, nous cherchons à examiner si l'effet

4. Dans l'échantillon considéré, ceci correspond à 177 catégories, à comparer avec les 487 postes de l'échelon le plus détaillé, à 4 caractères.

5. La définition de des communes couvertes et non-couvertes est la même que celle retenue dans le reste de ce chapitre.

mesuré est tiré par une période particulière de l'année 2020. Nous restreignons ainsi l'échantillon dans la colonne (4) aux entretiens ayant eu lieu entre le 18 mars 2020 et la fin du mois de juin. L'effet perd de sa significativité mais demeure positif.

Dans les colonnes (5) et (6), nous introduisons dans notre spécification des effets fixes au niveau communal, de telle sorte que ces effets fixes captent totalement les variations de télétravail dues aux caractéristiques communales, et en particulier le taux de couverture de la fibre au niveau communal. L'effet propre de la couverture fibre sur la propension au télétravail est ainsi neutralisé, et seul l'effet d'interaction entre couverture et télétravaillabilité des postes est identifié. Les colonnes (5) et (6) diffèrent uniquement par la finesse des effets fixes au niveau PCS que nous introduisons, la colonne (6) introduisant des effets fixes au niveau des catégories définies à 4 caractères, soit le niveau le plus fin possible. Les résultats demeurent largement inchangés. Enfin la colonne (7) restreint l'échantillon d'abord au deuxième trimestre 2020 pour examiner la spécificité de cette période particulièrement touchées par les restrictions sanitaire, sans qu'il soit possible de conclure à un effet différent que sur le reste de la période de crise pour laquelle les données sont disponibles (mars à décembre). La colonne (8) reproduit l'analyse sur l'année 2019. L'effet de l'accessibilité fibre est toujours d'autant plus positif que les occupations sont intenses en outils numériques, mais dans une moindre mesure.

TABLEAU 2.7 – Télétravail pendant la crise

	Teletravail							
	Mars-Déc. 2020			T2 2020	Mars-Déc. 2020		T2 2020	2019
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
part_info	0,523***	0,440***	0,195***	0,291***				
	0,013	0,015	0,034	0,071				
fibre	-0,016*	-0,024**	-0,019*	0,007				
	0,007	0,008	0,007	0,017				
1,fibre#c,part_info	0,190***	0,089***	0,061***	0,052	0,065***	0,096***	0,049	0,041***
	0,018	0,018	0,018	0,036	0,019	0,019	0,039	0,012
EF mois	O	O	O	O	O	O	O	O
Sexe & EF âge	N	O	O	O	O	O	O	O
EF secteurs A88	N	O	O	O	O	O	O	O
EF département	N	O	O	O	N	N	N	N
EF commune	N	N	N	O	O	O	O	O
EF PCS 3 caractères	N	N	O	O	O	N	N	N
EF PCS 4 caractères	N	N	N	N	N	O	O	O
N	22007	20975	20975	6044	20975	20975	6044	43454
r2	0,25	0,33	0,40	0,46	0,48	0,46	0,58	0,32

Notes : Les écarts-types robustes sont reportés entre parenthèses. + $p < .1$ * $p < 0.10$ ** $p < 0.01$, *** $p < 0.001$

Le tableau 2.8 utilise les mêmes spécifications mais examine l'effet de l'accessibilité de la fibre et la télétravaillabilité des postes sur le recours à l'activité partielle⁶. Elle met en lumière un effet d'interaction négatif entre disponibilité de la fibre et intensité en tâches numérique sur la probabilité de recourir à l'activité partielle. Par exemple, le coefficient de la colonne (3) implique que la couverture fibre est associée à une baisse de la probabilité du recours à l'activité partielle de 1,6 point de pourcentage. À nouveau, ces résultats sont robustes à l'introduction de variables contrôlant finement pour les caractéristiques individuelles et d'emploi.

Ce résultat, mis en regard avec ceux sur le recours au télétravail, illustre que la qualité des infrastructures numériques locales a pu avoir un effet important durant la crise sanitaire dans l'arbitrage auquel ont fait face les employeurs et leurs salariés, entre la poursuite de leur activité en télétravail et le recours à l'activité partielle.

6. Le nombre d'observations est beaucoup plus élevé dans ce tableau, car au contraire des questions sur le travail à domicile, qui ne sont posées qu'à 1/3 des répondants, l'Enquête Emploi en Continu collecte des informations sur l'activité partielle auprès de tous les répondants.

TABLEAU 2.8 – Activité partielle pendant la crise

	Part d'activité partielle								
	Mars-Déc. 2020			T2 2020		Mars-Déc. 2020		T2 2020	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
part_info	-0,049*** 0,005	-0,045*** 0,006	-0,032+ 0,017	-0,073+ 0,039	-0,037* 0,017				
fibre	0,033*** 0,005	0,012* 0,005	0,009+ 0,005	0,003 0,012					
1,fibre#c,part_info	-0,041*** 0,008	-0,023** 0,008	-0,016* 0,008	-0,005 0,019	-0,018* 0,008	-0,013 0,008	0,000 0,020	0,003 0,002	
EF mois	O	O	O	O	O	O	O	O	
Sexe & EF âge	N	O	O	O	O	O	O	O	
EF secteurs A88	N	O	O	O	O	O	O	O	
EF département	N	O	O	O	N	N	N	N	
EF commune	N	N	N	N	O	O	O	O	
EF PCS 3 caractères	N	N	N	O	O	N	N	N	
EF PCS 4 caractères	N	N	N	N	N	O	O	O	
N	58195	56384	56384	17870	56384	56384	17870	117000	
r2	0,13	0,20	0,21	0,21	0,25	0,26	0,33	0,19	

Notes : Les écarts-types robustes sont reportés entre parenthèses. + $p < .1$ * $p < 0.10$ ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

2.7 Conclusion

Dans ce chapitre, nous avons élaboré un ensemble de résultats concordants sur le rôle qu'ont pu jouer les infrastructures numériques dans la crise sanitaire. Celle-ci a mis à épreuve la résilience des entreprises et des économies locales.

En examinant les évolutions au niveau des communes, nous trouvons que la présence d'une couverture fibre au niveau de la commune a amorti les effets du choc en termes d'offres d'emploi, qui ont moins décliné dans les communes couvertes par la fibre que dans des communes comparables mais non-couvertes par la fibre ; et ceci s'est également traduit par un nombre de nouvelles embauches supérieur, bien qu'encore très négativement affecté par l'occurrence de la crise sanitaire. En partie grâce à la présence du dispositif d'activité partielle, l'emploi total au niveau communal a été moins affecté par la crise, et l'ampleur du choc sur l'emploi est donc moins corrélée avec la présence d'infrastructures numériques au niveau local.

L'examen de données individuelles confirme ces résultats agrégés et permet d'en explorer les mécanismes : le fait de résider dans des communes couvertes par la

fibre a favorisé le passage en télétravail, et a contrario, fait baisser le recours à l'activité partielle, des actifs occupés dans des professions reposant plus fortement sur les technologies numériques.

Enfin, le marché immobilier, dynamique pendant la crise sanitaire, a également vu une différence apparaître selon la couverture des localités en termes de fibre ; ainsi, les communes couvertes par la fibre ont vu le nombre de transactions immobilières augmenter par rapport aux communes non-couvertes.

2.8 Annexe

TABLEAU 2.9 – Statistique descriptives selon couverture au 2020T1 : ensemble des communes

	(1) Fibre = 0	(2) Fibre = 1	(3) Différence
<i>Couverture</i>			
Part couverture fibre	0.0786	0.9202	0.8415***
Commune zone RIP	0.7848	0.7586	-0.0261***
<i>Structure population</i>			
Part de cadres	0.0720	0.1003	0.0283***
Part d'agriculteurs	0.0360	0.0190	-0.0170***
Secteurs forte exposition crise	0.1639	0.1551	-0.0088*
Population 2017	2192	4937	2744***
Log Population 2017	6.7579	7.1921	0.4342***
<i>Type de commune</i>			
Catégorie de densité (1 à 4)	3.0375	2.6549	-0.3826***
Commune rurale	0.7198	0.5668	-0.1530***
Commune de montagne	0.1759	0.0968	-0.0791***
<i>Tendances 2018-2019</i>			
Moyenne $\Delta_{t,t-4} \ln$ (offres) 2019-2018	0.0309	0.0022	-0.0288**
N	18171	4019	

Notes : Ce tableau présente la moyenne de certaines variables mesurées entre 2017 et 2019 selon le degré de couverture fibre de chaque ville au premier trimestre de 2020. La variable binaire "Fibre" qui définit le groupe de traitement est égale à 1 lorsque la part de locaux d'une ville disposant d'un accès à la fibre est supérieure à 75%. La variable est sinon égale à 0.

2.8.1 Construction de l'index de télétravaillabilité : L'enquête Conditions de Travail 2019

Dans cette annexe, nous reprenons, à des fins de clarté et d'exhaustivité, des éléments de travaux développés par plusieurs membres de l'équipe dans le cadre de l'appel à projets "Les conséquences de la crise sanitaire sur le marché du travail" porté par la DARES.

L'enquête Conditions de Travail 2019 représente une source précieuse afin d'élaborer une mesure de télétravaillabilité dans le contexte français. Datant de 2019, elle offre une image précise et instantanée de l'état des pratiques de travail par métier juste avant la crise.

Cette enquête, la huitième vague d'une enquête historique menée par la DARES, a été collectée entre 2018 et mai 2019 et a interrogé 24 951 individus. Elle est représentative de l'ensemble des actifs en emploi en France métropolitaine et dans les cinq départements d'Outre-mer couverts.

Le questionnaire de l'enquête couvre de très nombreux aspects du travail des enquêtés. Ceux-ci ont traité, entre autres, à la place des outils numériques dans le temps de travail, aux équipements utilisés pendant l'exercice de l'activité professionnelle (protection, sécurité, etc), mais aussi à certaines tâches particulières (marcher, interagir avec le public en face à face, porter des objets lourds). Il est ainsi possible de construire un ensemble d'informations pertinentes pour la pratique du télétravail, sur le modèle de Dingel et Neiman (2020). De manière cruciale, l'enquête comprend également des données sur la pratique effective du télétravail, ce qui permet de valider la pertinence des mesures sélectionnées et de potentiellement suivre les évolutions des pratiques à travers la période de crise sanitaire par rapport à un état des lieux la précédant immédiatement.

Nous sélectionnons un ensemble (une dizaine en tout) de questions qui recoupent le champ des indicateurs utilisés par Dingel et Neiman. Ces questions,

résumées dans le tableau 2.10, portent par exemple sur l'utilisation de matériel informatique, l'existence de tensions, le fait de travailler à l'extérieur, d'être exposé à divers risques professionnels, de devoir marcher, de faire face à du public, ou encore de porter des charges lourdes dans le cadre de son travail.

TABLEAU 2.10 – Extrait du questionnaire Conditions de Travail 2019

Nom de la variable	Intitulé de la question
part_info	Au total, pendant combien de temps utilisez-vous à titre professionnel les matériels informatiques dont nous venons de parler ?
tension	Vivez-vous des situations de tension dans vos rapports avec le public ? (usagers, patients, élèves, voyageurs, clients, fournisseurs, etc.) ?
exterieur	Où passez-vous la plus grande partie de votre temps de travail ? (réponse pertinente : en extérieur / chantiers)
securite	A votre emplacement de travail, êtes-vous amené à
securite1	... respirer des fumées ou des poussières ?
securite2	... être en contact avec des produits dangereux ?
securite3	... être exposé à des risques infectieux ?
securite4	... risquer d'être blessé ou accidenté ?
marcher	L'exécution de votre travail vous impose-t-elle d'effectuer des déplacements à pieds longs ou fréquents ?
protection	Votre employeur met-il à votre disposition des équipements individuels de protection comme des gants des lunettes des chaussures de sécurité, un harnais ? Les utilisez-vous ?
lourd	L'exécution de votre travail vous impose-t-elle de porter ou déplacer des charges lourdes ?
machine	Votre rythme de travail vous est-il imposé par la cadence automatique d'une machine ?
vehicule	Utilisez-vous un véhicule dans le cadre de votre travail ou pour vos besoins professionnels, en dehors des trajets domicile-travail ?
public	Etes-vous en contact direct avec le public ? De vive voix en face à face ?

Dans la perspective de comprendre les contraintes qui ont poussé certaines entreprises à avoir recours à l'activité partielle, ce premier chapitre propose une mesure de la « télétravaillabilité » des différentes activités productives. Pour cela, nous adoptons une approche par métier. Après avoir discuté la pertinence des mesures internationales existantes, nous proposons une nouvelle approche pour la France

basée sur l'exploitation de l'enquête Conditions de Travail 2019.

Nous commençons par examiner la corrélation entre les différents indicateurs et la pratique du télétravail au niveau individuel. Nos analyses révèlent qu'à l'exception des situations tensions, toutes les variables de condition de travail retenues sont significativement reliées tant au recours au télétravail, qu'à l'intensité de son utilisation (tableau 2.11).

Notre objectif étant de mieux appréhender les métiers pour lesquels le télétravail est possible, nous avons agrégé l'information disponible en prenant les moyennes de nos différentes variables d'intérêt au niveau de chaque catégorie socioprofessionnelle (CSP) détaillée. Deux variables en particulier, la part du temps de travail passé sur des outils numériques et l'utilisation d'équipements de protection, s'avèrent posséder un pouvoir explicatif important dans l'explication du télétravail en 2019. Ceci est illustré par les figures 2.4 et 2.5. La figure 2.4 montre que la part de salariés télétravaillant au moins une heure par mois dans une CSP détaillée est positivement corrélée à la part du temps de travail moyenne pour laquelle les salariés déclarent utiliser des outils numériques. La figure 2.5 montre une corrélation cette fois négative du recours au télétravail en 2019 avec la part de salariés de la CSP déclarant porter des équipements de protection ⁷.

On remarque sur ces figures que la relation entre télétravail et recours aux outils numériques ou aux équipements de protection au sein d'un métier n'est pas continue, et qu'il existe une forte variabilité des configurations. Par exemple, pour un certain nombre de métiers ayant très fortement recours aux outils numériques, le télétravail reste très peu répandu en 2019. En revanche, on voit que pour pratiquement tous les métiers pour lesquels le recours aux outils numériques est faible (moins de 20% du temps de travail des salariés en moyenne), le télétravail est quasiment inexistant. Cette distinction illustre la logique d'une approche par les conditions de

7. Sur ces figures, chaque point représente une CSP et la taille du point est proportionnelle au nombre d'observations disponibles dans la CSP dans l'enquête CT 2019.

FIGURE 2.4 – Temps passé sur les outils numériques et probabilité de télétravail

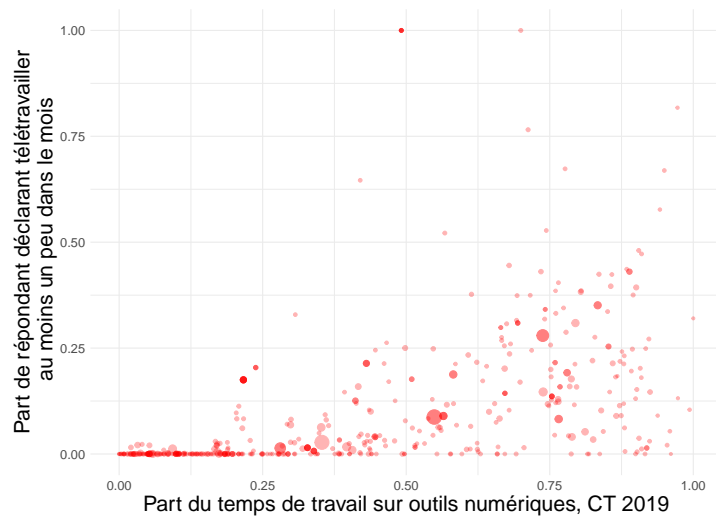
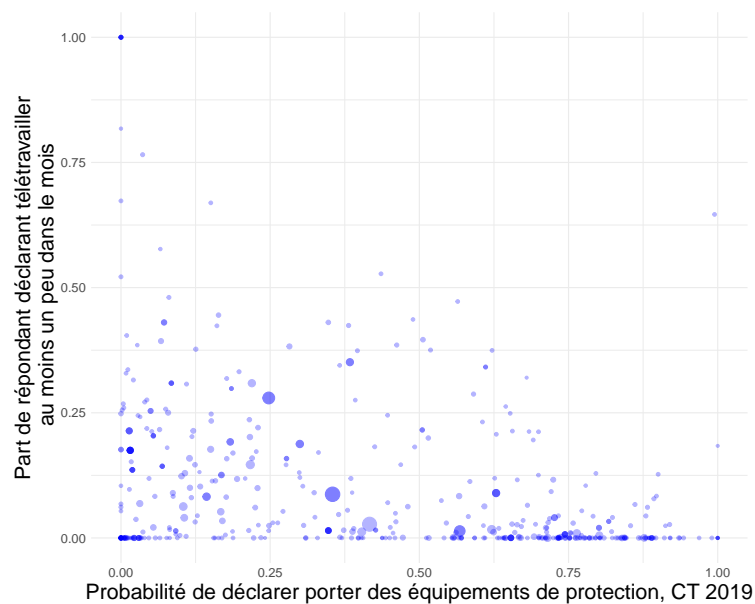


FIGURE 2.5 – Probabilité de porter des équipement de protection dans l'exercice de sa profession et télétravail



travail plutôt que par le recours antérieur au télétravail pour prédire la capacité à télétravailler pendant la crise Covid : le recours aux outils numériques semble une condition sine qua none du fait de pouvoir télétravailler, et pour autant, les métiers ayant recours aux outils numériques avant la crise n'utilisaient pas nécessairement le télétravail.

Le même type d'enseignement peut être tiré de la figure 2.5 qui montre que les métiers requérant de porter tout le temps ou presque des équipements de protection semblent très peu télétravaillables, tandis que la réciproque n'est pas vraie.

TABLEAU 2.11 – Coefficients de différents régresseurs dans une régression linéaire univariée avec deux mesures de télétravail en 2019

Variable dépendante	Part de télétravail	SE	Télétravail >0	SE
Régresseur				
part_info	0,06	0,01	0,29	0,02
tension	0,01	0,02	0,02	0,04
exterieur	-0,04	0,01	-0,15	0,04
securite	-0,05	0,01	-0,25	0,02
securite1	-0,04	0,01	-0,22	0,02
securite2	-0,06	0,01	-0,24	0,03
securite3	-0,06	0,01	-0,25	0,02
securite4	-0,06	0,01	-0,26	0,03
marcher	-0,08	0,01	-0,35	0,03
protection	-0,06	0,01	-0,19	0,02
lourd	-0,04	0,01	-0,24	0,02
vehicule	0,03	0,01	0,08	0,03
public	-0,01	0,01	-0,09	0,02

Note : Les coefficients dans les deuxième et quatrième colonnes représentent les coefficients de chaque variable spécifiée dans la première colonne, dans une régression univariée avec la part de télétravail (deuxième colonne) et une variable indicatrice qui prend 1 comme valeur lorsque cette part est non nulle. Les valeurs dans les troisièmes et cinquièmes colonnes correspondent aux erreurs standards.

2.8.2 Tests de robustesse

TABLEAU 2.12 – Évolution des embauches 2020T2 : Davis-Haltiwanger

	DH(embauches) T2 2019 2020			— 2019-2018
	(1)	(2)	(3)	(4)
Forte Couverture Fibre 2020T1	0,0604** (0,0257)	0,0648** (0,0256)	0,0697*** (0,0257)	0,000482 (0,0249)
Part de cadres		-0,491*** (0,115)	-0,424*** (0,118)	0,0182 (0,115)
Part d'agriculteurs		0,236 (0,160)	-0,0445 (0,185)	-0,324** (0,158)
Secteurs forte exposition crise		-0,324*** (0,0311)	-0,340*** (0,0316)	0,0927*** (0,0267)
R^2	0,000	0,012	0,014	0,001
Adjusted R^2	0,000	0,012	0,013	0,001
Effet fixe densité + rural + mont.			✓	✓
Placebo				✓
Observations	13852	13397	13397	15126

Notes : Cette table présente les résultats d'une régression en première différence (voir équation (3.1) pour plus de détails). La variable dépendante est le taux de croissance Davis Haltiwanger du nombre d'embauches entre 2020T2 et 2019T2. La colonne (4) présente un test de pré-tendance commune entre utilisant le taux de croissance entre 2019T2 et 2018T2 comme variable dépendante. L'échantillon d'estimation inclut l'ensemble des communes n'ayant aucune couverture fibre en 2019T1. Les communes définies comme ayant une forte couverture au moment de la crise Covid sont celles ayant une couverture supérieure à 75% au premier trimestre de 2020 selon les données de l'Arcep sur le déploiement de la fibre. Les variables de contrôles sont les suivantes : *Secteurs forte exposition crise* calcule la part de l'emploi dans des secteurs fortement affectés par la crise en 2019. Parmi les effets fixes : *densité* fait référence aux catégories de densité définies par l'Insee (4 valeurs), rural et montagne sont des variables binaires. Les contrôles démographiques (CSP) contrôle pour la part de cadre et de retraités dans la population en 2017-2018. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLEAU 2.13 – Offres d'emploi 2019T2-2020T2 : groupe de contrôle restreint.

	$\Delta_{t,t-4} \ln$ (offres) T2 2019 2020			— 2019-2018
	(1)	(2)	(3)	(4)
Forte Couverture Fibre 2020T1	0,207*** (0,0440)	0,179*** (0,0438)	0,157*** (0,0458)	-0,0474 (0,0415)
Part de cadres		-0,464 (0,312)	-0,337 (0,319)	0,0762 (0,316)
Part d'agriculteurs		3,338*** (0,689)	2,686*** (0,794)	-2,522*** (0,758)
Secteurs forte exposition crise		-0,508*** (0,103)	-0,480*** (0,105)	0,0752 (0,0949)
R^2	0,012	0,040	0,044	0,009
Adjusted R^2	0,012	0,038	0,040	0,005
Effet fixe densité + rural + mont, Placebo			✓	✓
Observations	2094	2090	2090	2243

Notes : La variable dépendante est le taux de croissance du nombre d'offres d'emploi entre 2020T2 et 2019T2. La colonne (4) présente un test de pré-tendance commune entre utilisant le taux de croissance entre 2019T2 et 2018T2 comme variable dépendante. L'échantillon d'estimation inclut l'ensemble des communes n'ayant aucune couverture fibre en 2019T1 et qui ont connu une **hausse strictement positive de leur couverture au cours de l'année 2019**. Les communes définies comme ayant une forte couverture au moment de la crise Covid sont celles ayant une couverture supérieure à 75% au premier trimestre de 2020 selon les données de l'Arcep sur le déploiement de la fibre. Les variables de contrôles sont les suivantes : *Secteurs forte exposition crise* calcule la part de l'emploi dans des secteurs fortement affectés par la crise en 2019. Parmi les effets fixes : *densité* fait référence aux catégories de densité définies par l'Insee (4 valeurs), rural et montagne sont des variables binaires. Les contrôles démographiques (CSP) contrôlent pour la part de cadre et de retraités dans la population en 2017-2018. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLEAU 2.14 – Embauches 2019T2-2020T2 : groupe de contrôle restreint

	$\Delta \ln(\text{embauches})_{t,t-4}$		T2 2019-20	— 2019-2018
	(1)	(2)	(3)	(4)
Forte Couverture Fibre 2020T1	0,118*** (0,0367)	0,100*** (0,0363)	0,111*** (0,0377)	0,0145 (0,0448)
Part de cadres		-0,525** (0,261)	-0,553** (0,265)	0,447 (0,307)
Part d'agriculteurs		0,382 (0,684)	0,248 (0,803)	0,251 (0,693)
Secteurs forte exposition crise		-0,444*** (0,106)	-0,463*** (0,108)	0,190* (0,103)
R^2	0,004	0,015	0,018	0,003
Adjusted R^2	0,004	0,014	0,015	-0,000
Effet fixe densité + rural + mont.			✓	✓
Placebo				✓
Observations	2805	2746	2746	2968

Notes : La variable dépendante est le taux de croissance du nombre d'embauches entre 2020T2 et 2019T2. La colonne (4) présente un test de pré-tendance commune entre utilisant le taux de croissance entre 2019T2 et 2018T2 comme variable dépendante. L'échantillon d'estimation inclut l'ensemble des communes n'ayant aucune couverture fibre en 2019T1 et qui ont connu une **hausse strictement positive de leur couverture au cours de l'année 2019**. Les communes définies comme ayant une forte couverture au moment de la crise Covid sont celles ayant une couverture supérieure à 75% au premier trimestre de 2020 selon les données de l'Arcep sur le déploiement de la fibre. Les variables de contrôles sont les suivantes : *Secteurs forte exposition crise* calcule la part de l'emploi dans des secteurs fortement affectés par la crise en 2019. Parmi les effets fixes : *densité* fait référence aux catégories de densité définies par l'Insee (4 valeurs), rural et montagne sont des variables binaires. Les contrôles démographiques (CSP) contrôlent pour la part de cadre et de retraités dans la population en 2017-2018. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

CHAPITRE 3

ANALYSE DES EFFETS DU DÉPLOIEMENT DE LA FIBRE AU NIVEAU LOCAL

Dans ce chapitre, nous développons une analyse à l'échelle très locale, où l'unité de base est la parcelle cadastrale. Nous tirons parti de la très grande finesse et précision des données Arcep sur le déploiement de la fibre dans chaque immeuble de France, pour examiner les évolutions de la démographie, des revenus des ménages, et du marché immobilier.

3.1 Données

Données de connexion à l'immeuble (Arcep). Les données Arcep à l'échelle des immeubles contiennent des informations sur les dates de mise en service du point de mutualisation (PM) et du point de branchement optique (PBO) correspondant à chaque immeuble. Chaque immeuble identifié dans les données est par ailleurs géolocalisé par ses coordonnées géographiques, que nous utilisons pour les associer à une parcelle cadastrale du territoire. Pour ce faire, nous apparions chaque immeuble avec la parcelle cadastrale bâtie qui le contient, ou, s'il n'est contenu dans le périmètre d'aucune, à la plus proche d'entre elles, dans la limite de 4 mètres de

distance (ceci apparie la quasi-totalité des immeubles de la base)¹. Si une parcelle cadastrale se voit attribuer plusieurs immeubles, nous lui allouons les dates les plus anciennes de raccordement PM et PBO.

Le statut de traitement privilégié que nous retenons est la date (année ou trimestre) de raccordement du point de mutualisation, qui est susceptible d'être davantage décorrélée des évolutions passés des caractéristiques économiques ou démographiques de la parcelle. La date de mise à disposition du point de branchement optique est certes une mesure plus précise du statut de raccordement spécifique de l'immeuble, mais elle est également plus probablement endogène à des évolutions récentes au sein de la parcelle, car elle est davantage fonction de la demande des occupants d'une parcelle.

Données Fidéli (Insee et DGFIP). Le Fichier démographique d'origine fiscale sur les logements et les individus (Fidéli) est constitué d'un appariement entre des données fiscales (principalement issues de la taxe d'habitation et de l'impôt sur le revenu), associées à des informations sur la localisation précise des logements (parcelle cadastrale) et leurs caractéristiques. Si les millésimes de Fidéli ne peuvent pas être liés entre eux pour construire un panel d'individus, chaque millésime contient les informations de revenus et de localisation pour l'année N et pour l'année N-1. Cette caractéristique permettrait normalement de décomposer au niveau de chaque parcelle les variations de revenus issues de déménagement, cependant des problèmes affectant certains millésimes de la base² ne nous permettent pas d'adopter cette approche. Nous utilisons les millésimes 2016, 2017, 2019 et 2020 du fichier pour construire un panel au niveau parcelle cadastrale du nombre d'habitants et de

1. En effet, après examen des cartes cadastrales et Arcep, il s'avère que l'enregistrement des coordonnées géographiques d'un immeuble se fait de façon régulière au niveau des boîtiers de raccordement, qui, en zone urbaine, peuvent se trouver dans les rues et voies, ce qui impose une certaine souplesse dans l'appariement géographique.

2. Millésimes 2015 (première année de production) et 2018 (2 millions d'occurrences fiscales TH manquantes), voir comptes-rendus des comités d'utilisateurs de Fidéli.

leurs revenus sur les années 2014 à 2019³

Fichiers fonciers (DGFIP). Les données des fichiers fonciers, disponibles pour les millésimes 2013 à 2019, sont issues de l'enrichissement du cadastre avec des données d'origine fiscale. Ils couvrent ainsi l'ensemble du territoire français, ainsi que la plupart des immeubles et locaux (à l'exception de nombreux bâtiments publics). Ils permettent de disposer, de façon annuelle, du statut d'occupation ou de vacance des immeubles, ainsi que de certaines informations sur les caractéristiques de leur occupation (nature résidentielle ou commerciale, secteur d'activité, etc.).

DV3F (Cerema). Les données DV3F sont issues du croisement des données des Demandes de Valeurs Foncières (DGFIP), contenant l'ensemble des mutations foncières, avec les Fichiers Fonciers.

3.2 Approche empirique

3.2.1 Étude d'évènement échelonnée

Une approche par étude d'évènements. L'étude d'évènements, ou en anglais *event study*, est une méthodologie usuelle de l'évaluation d'impact. Elle exploite une multiplicité d'évènements dont la temporalité est considérée comme exogène. En estimant les effets temporels avant et après l'évènement, cette méthode vise généralement à obtenir une mesure causale du traitement (l'évènement) sur des variables d'intérêt, qui peuvent être les revenus au sein d'une parcelle cadastrale, ou la probabilité qu'une transaction ait lieu dans cette parcelle. Cette méthode est très proche de la différence-de-différences, la différence entre ces deux méthodes résidant dans le fait que les unités reçoivent dans le premier cas le traitement à des

3. La structure des fichiers est telle que, par exemple, le millésime 2020 contient les revenus déclarés en 2020 (année N) et 2019 (année N-1), ce qui correspond en réalité à des revenus perçus en 2019 (N) et 2018 (N-1).

dates différentes. Dans le cas que nous étudions, l'ensemble des unités considérées dans l'estimation sont finalement traitées.

Le statut d'exogénéité de la date de connexion au THD d'une parcelle est une hypothèse discutable, puisqu'il est possible que le choix de connecter une parcelle plutôt qu'une autre soit corrélé avec le nombre d'habitants ou les revenus de ses habitants. À ce titre, l'analyse des tendances pré-traitement (coefficients de régression associés aux années précédant la connexion THD) donnera une première mesure permettant de juger de la plausibilité de cette hypothèse d'exogénéité.

Spécification. Nous estimons une spécification dynamique qui autorise l'effet du traitement à varier en fonction du temps écoulé depuis le traitement. L'année de traitement d'une entreprise i est notée t_{i0} . On peut alors indexer le temps écoulé par rapport au traitement avec $d = t - t_{i0}$ où t est une année calendaire. L'indice d est négatif avant le traitement et positif après. La fenêtre d'estimation est définie par le biais d'une borne inférieure m_0 et supérieure m_1 , qui varient en fonction des données sur lesquelles l'estimation est menée. Les observations incluses dans l'échantillon d'estimation sont donc telles que : $d \in \{m_0, m_0 + 1, \dots, m_1 - 1, m_1\}$.

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

$$Y_{it} = \sum_{\substack{d=m_1 \\ d=m_0+1 \\ d \neq -1}} \beta_d \times \mathbf{1}\{t = d + t_{i0}\} + \mathbf{x}'_{it} \boldsymbol{\delta} + \alpha_i + \psi_t + \varepsilon_{it}, \quad (3.1)$$

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

On propose ici d'omettre la variable indicatrice pour $d = -1$ ainsi que la variable indicatrice pour $d = m_0$. Cette restriction est nécessaire dans le cas d'une étude d'évènements comme celle que nous menons où l'ensemble des unités considérées sont finalement traitées, pour éviter la multi-collinéarité et pour identifier le processus de génération de données dynamiques sous-jacent si l'ensemble des unités

incluses dans l'échantillon d'estimation est traité (Borusyak et Jaravel, 2017; Gross et al., 2019).

3.2.2 Étude d'évènement empilée avec sélection explicite du groupe de contrôle

Constitution des groupes traité et contrôle selon des critères géographiques.

La méthode par étude d'évènement échelonnée est proche dans l'esprit de l'approche adoptée dans le premier chapitre de ce rapport. Nous développons dans cette section une approche locale qui exploite plus finement et plus directement le degré élevé de précision géographique des données au niveau infra-communal. Cette méthode vise à sélectionner de façon explicite un groupe de contrôle qui est géographiquement proche des unités traitées (c'est-à-dire connectées à la fibre)⁴.

Nous procédons de la façon suivante. Nous classons les parcelles cadastrales par date de connexion. Nous définissons une parcelle comme connectée si elle comprend ou bien se situe à moins de 4 mètres d'un point de mutualisation (PM) qui a été équipé à la fibre selon la base immeuble de l'Arcep. Nous définissons ensuite comme cohorte l'ensemble des parcelles ayant été connecté à la même période t (une année).

Pour chaque cohorte de raccordement, on vise à identifier un ensemble de parcelles traitées et un ensemble de parcelles contrôles. À la date t , les parcelles traitées sont les parcelles raccordées à la date t et possédant au moins une parcelle contrôle.

Pour une date t , les parcelles contrôles sont les parcelles non encore raccordées

4. Une approche envisagée initialement avait de mettre en place une régression de discontinuité spatiale. Néanmoins, cette approche est rendue difficile par le fait qu'il n'existe pas de frontière explicite dans notre cas, et que le caractère incrémental du déploiement rend difficile la détermination de la frontière pertinente. En l'espèce, notre stratégie se rapproche d'une différence de discontinuité non-paramétrique : à mesure que le rayon du périmètre considéré pour la sélection des unités traitées et des unités contrôles tend vers 0, les unités sélectionnées se rapprochent de la frontière géographique du déploiement du traitement. L'approche que nous retenons ici présente l'avantage considérable de court-circuiter le besoin d'estimer le périmètre effectif de la frontière du déploiement, tout en élaborant un estimateur à la définition et aux propriétés très similaires.

dont la date de raccordement est ultérieure à $t + 2$ an situées à *proximité* d'au moins une parcelle raccordée. Le rayon qui détermine la distance maximale pour définir *proximité* varie entre 100m et 1km selon la densité du bâti autour de la parcelle raccordée, où la densité est calculé sur la base d'un rayon de 5km. Plus précisément, notons d_p la densité autour de la parcelle p , notons d_{p1} et d_{p99} les centiles 1 et 99 de la distribution de la densité, la distance maximale pour qu'une parcelle soit considérée comme un contrôle potentiel est alors donnée par la formule suivante :

$$\text{dist}^{\max}(d_p) = 100 \times \mathbb{1}\{d_p \geq d_{p99}\} + 1000 \times \mathbb{1}\{d_p \leq d_{p1}\} + \left(100 + \frac{100 - 1000}{d_{p99}} \times d_p\right) \times \mathbb{1}\{d_p \in (d_{p1}, d_{p99})\} \quad (3.2)$$

La figure 3.1 présente les communes de France métropolitaine qui contiennent des parcelles traitées et contrôles pour la cohorte de parcelles raccordées au cours du premier trimestre 2017. Nous présentons ci-dessous un zoom sur la région du Mans (Figure 3.2) en prenant l'exemple de deux communes (Figure 3.3, panel a). Le panel (b) montre les différentes parcelles bâties qui sont traitées (vertes), contrôles (rouges) et écartées de l'analyse pour cette cohorte (grisées).

Dans le tableau 3.1, nous présentons des statistiques descriptives concernant les ménages au sein des groupes traitement et contrôle au sein de la base épi-lée. Ces statistiques portent ici sur l'ensemble des cohortes. Les deux premières colonnes présente la moyenne et les écarts type au sein du groupe de traitement à $t - 1$ (c'est-à-dire 1 an avant la connexion de la parcelle). Les colonnes 3 et 4 présentent les mêmes statistiques concernant le groupe contrôle. Enfin la colonne 5 présente la différence entre les moyennes des deux groupes et la colonne 6 reporte la p-value associée à ces différences de moyennes. Cette sixième colonne indique clairement que dans la plus part des cas on peut rejeter statistiquement l'égalité de moyenne entre les deux groupes. Néanmoins d'un point de vue quantitatif, ces différences apparaissent modérées voir faibles. Par exemple, si l'on considère les

TABLEAU 3.1 – Statistiques descriptives au niveau parcelle cadastrale pour l’analyse par étude d’évènement empilée

	Groupe traitement		Groupe contrôle		Différence	
	Moyenne	écart-type	Moyenne	écart-type	Diff.	p-value
Carac. parcelles en t-1						
Nombre d’adultes	4,574	20,952	4,477	16,537	-0,097	(0,004)
Nombre de foyers fiscaux	3,474	16,256	3,439	13,047	-0,035	(0,180)
Âge moyen des adultes	55,638	15,673	54,965	15,755	-0,673	(0,000)
Moyenne des revenus d’activité	15350,730	21498,764	15050,557	22092,922	-300,173	(0,000)
Moyenne des salaires	14511,991	19872,658	14187,246	20900,223	-324,745	(0,000)
Moyenne des BIC	324,022	4271,334	330,167	3998,801	6,145	(0,407)
Moyenne des retraites	8179,620	11181,236	7848,241	11750,982	-331,379	(0,000)
Moyenne des alloc, chômage	603,638	2202,399	638,373	2244,576	34,735	(0,000)
P(Salaires > 0)	0,671	0,470	0,683	0,465	0,012	(0,000)
P(Retraites > 0)	0,528	0,499	0,523	0,499	-0,006	(0,000)
P(Alloc, chômage > 0)	0,216	0,412	0,233	0,423	0,017	(0,000)
P(BIC > 0)	0,054	0,226	0,057	0,232	0,003	(0,000)
Carac. communes						
AMII	0,765	0,424	0,767	0,422	0,002	(0,004)
RIP	0,091	0,288	0,082	0,275	-0,009	(0,000)
Commune rurale	0,034	0,182	0,023	0,149	-0,012	(0,000)
Commune de montagne	0,070	0,255	0,048	0,213	-0,022	(0,000)
Observations	606887		655102		1261989	

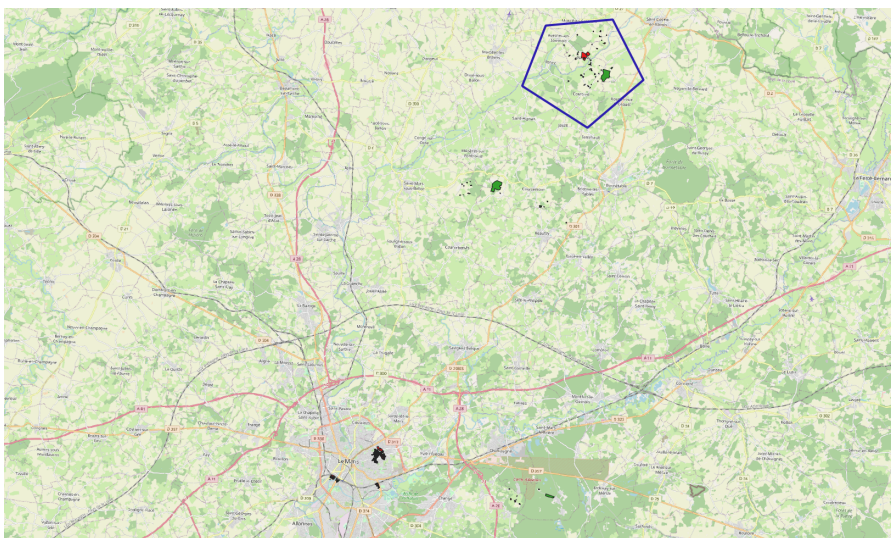
revenus d’activités, on constate que les ménages au sein des parcelles traitées ont un revenu moyen inférieur aux parcelles traitées, mais cette différence est de 300 euros, ce qui représente un montant faible relativement au revenu moyen (15 000 euros) ou a fortiori à la dispersion de cette variable (avec un écart type d’environ 21 500 euros). Une grande limite de cette approche empilée cependant est de n’inclure qu’une part limitée de communes RIP : environ 9 % des parcelles analysées sont dans ces communes RIP contre 34 % dans l’échantillon complet utilisé dans le cadre de l’approche échelonnée (voir infra le Tableau 3.2).

FIGURE 3.1 – Les traités de 2017 T1 et leurs contrôles : France métropolitaine



Notes : Cette figure présente les traités de la cohorte 2017 T1 et leurs contrôles, au sein de la France métropolitaine.

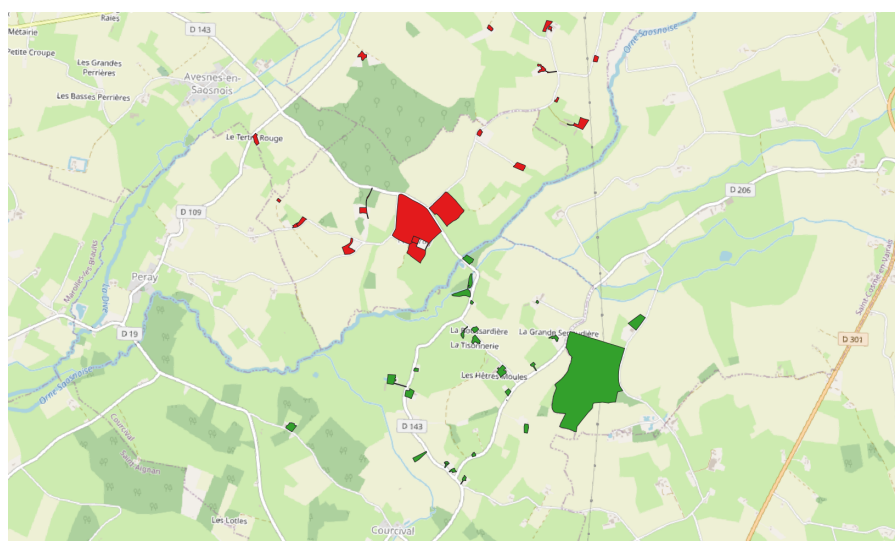
FIGURE 3.2 – Les traités de 2017 T1 et leurs contrôles : exemple dans la région du Mans



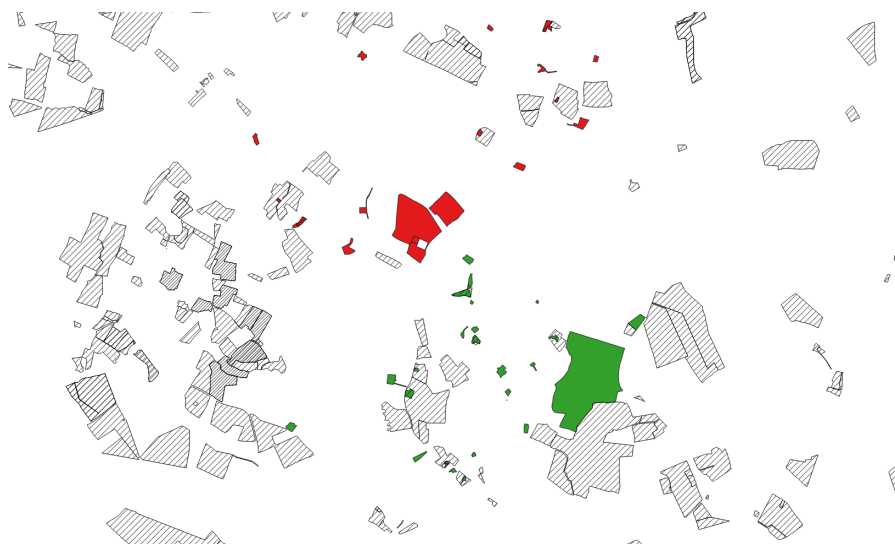
Notes : Cette figure présente les traités de la cohorte 2017 T1 et leurs contrôles au sein de communes voisines du Mans.

FIGURE 3.3 – Les traités de 2017 T1 et leurs contrôles : exemple dans la région du Mans

(a) Parcelles traitées et contrôles : avec fond de carte.



(b) Parcelles traitées et contrôles : sans fond de carte.



Notes : Cette figure présente les traités de la cohorte 2017 T1 et leurs contrôles, au sein de la région du Mans.

Création de la base d'estimation empilée. Nous adoptons la méthode d'étude d'évènement empilée, employée par Vannutelli (2020), qui est également similaire à Cengiz et al. (2019) et Deshpande et Li (2019).⁵ Cette dernière consiste à construire un *groupe contrôle* pour chaque cohorte traitée. En pratique, pour chaque cohorte de parcelles entre 2013 et 2020, nous construisons un échantillon séparé où nous définissons le temps de traitement relatif à cette cohorte spécifique, et où les parcelles contrôles (qui appartiennent pour certaines à des cohortes ultérieures) sont sélectionnées selon la procédure décrite dans le paragraphe précédent. Les observations au sein des cohortes de contrôle ne sont considérées que sur la période précédant leur traitement respectif. Finalement, nous empilons l'ensemble des échantillons (cohortes 2013 à 2019) et obtenons ainsi notre base d'estimation finale. La spécification que nous estimons et présentons dans le paragraphe ci-dessous prend la structure empilée de la base d'estimation en compte.

Spécification. On estime ensuite le modèle suivant :

$$Y_{pt} = \sum_{\substack{l=f_0 \\ l \neq -1}}^{l_0} \mathbb{1}\{t - T_0(p) = l\} \times T_p \times \beta_l + \delta_{d(p), T_0(p), t} + \delta_p + \delta_l + \alpha_p + \varepsilon_{pt} \quad (3.3)$$

où

- t est la période "absolue" (l'année calendaire), l est la période relative au traitement.
- p est une combinaison unique de parcelle-base d'estimation où une base d'estimation correspond à l'ensemble des traitées et contrôles pour une cohorte donnée⁶.
- $T_p = 1$ si la parcelle p fait partie du groupe de traitement.

5. Baker et al. (2021) montrent que la méthode des études d'évènement empilées donne des résultats similaires à la méthode proposée par Callaway et Sant'Anna (2020).

6. En tant que traitée, chaque parcelle n'est associée qu'à une seule cohorte. Une même parcelle peut cependant servir de contrôle pour une cohorte antérieure. Par exemple, une parcelle connectée à la fibre en 2018 apparaîtra comme traitée au sein de la base d'estimation 2018 et peut servir de contrôle pour la base d'estimation 2015.

- $T_0(p)$ est la cohorte/base d'estimation de la parcelle p .
- Y_{pt} une variable d'intérêt au niveau de la parcelle et de l'année
- $\delta_{d(p),t,T_0(p)}$ effet fixe année \times aire géographique à laquelle appartient la parcelle p , cet effet fixe est spécifique à chaque cohorte $T_0(p)$; α_p un effet fixe parcelle;
- f_0 et l_0 sont les premières et dernières années "relatives" observées pour la parcelle p pour chaque ensemble de parcelles définies par une même valeur de $T_0(p)$.

3.3 Impact de la connexion au THD sur la population et les revenus

Afin d'estimer l'impact de la connexion FTTH sur le nombre d'habitants et les revenus de ces habitants d'une parcelle cadastrale, nous procédons en première approche à une estimation par étude d'évènements échelonnée, telle que décrite dans la section 3.2.1. Afin de s'abstraire de chocs affectant les revenus dans certaines zones géographiques particulières au cours de la période, nous incluons des effets fixes "département \times année", en plus des effets fixes pour chaque parcelle cadastrale, dans nos régressions. En outre, les erreurs-standard sont groupées au niveau de la parcelle cadastrale.

L'échantillon d'estimation est composé de l'ensemble des parcelles cadastrales ayant au moins un habitant dans Fidéli chaque année de la période 2014–2019 (panel cylindré de parcelles cadastrales habitées). Nous construisons des variables de nombre d'individus (redevables fiscaux), moyenne des revenus d'activité (salaires, bénéfices industriels et commerciaux, bénéfices non commerciaux et bénéfices agricoles) dans la parcelle, moyenne des salaires, ainsi que des revenus de remplacement (allocations chômage, pensions de retraite), et l'âge moyen des redevables fiscaux dans la parcelle. Des variables dépendantes complémentaires sont égale-

ment la probabilité de déclarer des bénéfices industriels et commerciaux non-nuls, et la probabilité de déclarer des allocations chômage. La première probabilité nous renseigne sur la validité de l'hypothèse selon laquelle la connexion au THD aurait permis de développer des activités d'indépendants exigeantes en bande passante (par exemple des activités de développeurs, graphistes, etc). La seconde probabilité complète l'analyse réalisée dans le chapitre 2 sur l'interaction entre disponibilité de la fibre et chômage.

Afin d'approcher plus finement les effets induits par la connexion au THD dans les endroits dans lesquels l'action publique a été nécessaire, nous produisons l'ensemble des régressions en se restreignant aux parcelles contenues dans des zones RIP, en plus des estimations menées sur la France entière. Si les estimations perdent naturellement en précision du fait d'un échantillon plus restreint, elles renseignent sur un effet causal de la politique publique, puisqu'elles impliquent un effet causal de la connexion dans des zones dans lesquelles la connexion n'aurait a priori pas eu lieu sans soutien public.

Le tableau 3.2 présente des statistiques descriptives sur les principales variables de l'échantillon d'estimation, l'année précédant la connexion au THD de la parcelle. On observe que les parcelles considérées sont relativement peu denses (en moyenne 3 adultes répartis entre 2,2 foyers fiscaux). Ils reçoivent des revenus d'activité (somme des salaires, BIC, BNC et BA) atteint une moyenne d'environ € 15000, parmi lesquels la part des salaires est prépondérante (en moyenne € 14500). 64,6 % des parcelles incluent un redevable percevant des salaires, tandis que 51,8 % d'entre elles ont un redevable percevant des retraites, et 18,5 % d'entre elles ont un redevable bénéficiant d'allocations chômage. Les BIC quant à eux ne sont présents que dans 4,5 % des parcelles incluses dans l'échantillon.

Impact de la connexion au THD sur la population et l'âge des occupants Dans un premier temps, nous analysons l'impact de la connexion au THD sur la popula-

TABLEAU 3.2 – Statistiques descriptives au niveau parcelle cadastrale, échantillon complet

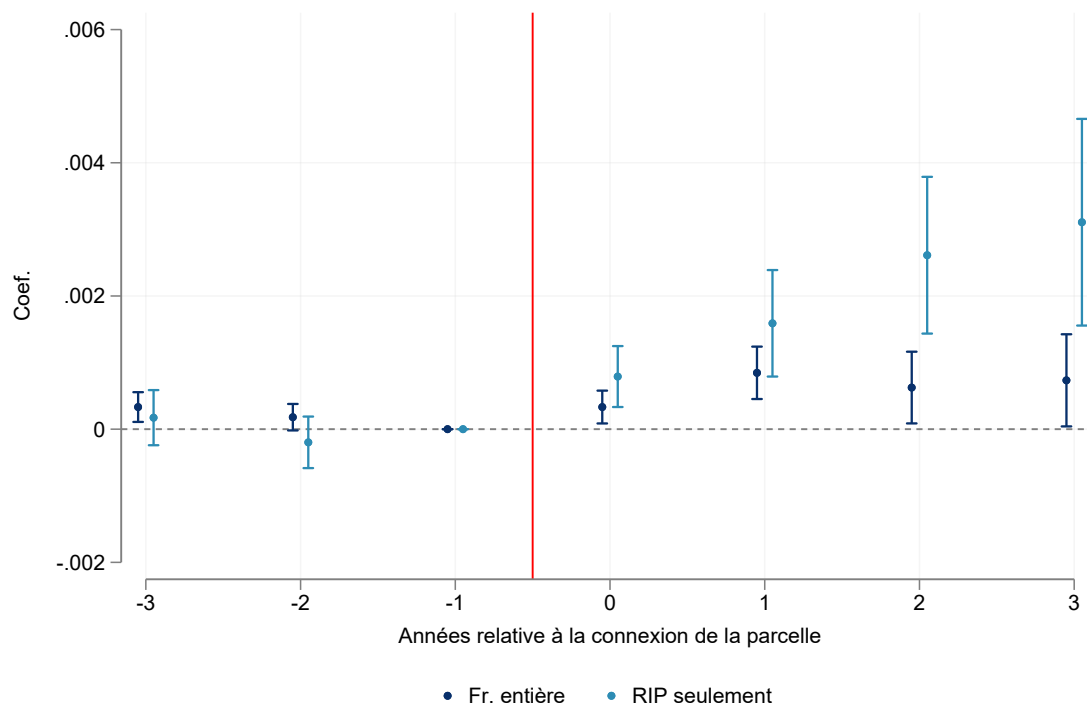
	Moyenne	écart-type	1er quartile	Médiane	3e quartile	N. obs.
Carac. parcelles en t-1						
Nombre d'adultes	3,080	12,065	2,000	2,000	2,000	6473093
Nombre de foyers fiscaux	2,256	9,483	1,000	1,000	2,000	6473093
Âge moyen des adultes	56,685	15,828	44,000	55,333	69,000	6473093
Moyenne des revenus d'activité	14951,767	20976,989	0,000	10661,750	23321,666	6473093
Moyenne des salaires	14588,034	22239,717	0,000	10452,000	23035,000	6405177
Moyenne des BIC	343,288	4297,604	0,000	0,000	0,000	6405177
Moyenne des retraites	8464,160	11270,622	0,000	2137,714	15274,000	6405177
Moyenne des alloc, chômage	583,826	2220,627	0,000	0,000	0,000	6405177
P(Salaires > 0)	0,646	0,478	0,000	1,000	1,000	6473093
P(Retraites > 0)	0,518	0,500	0,000	1,000	1,000	6473093
P(Alloc, chômage > 0)	0,185	0,388	0,000	0,000	0,000	6473093
P(BIC > 0)	0,045	0,207	0,000	0,000	0,000	6473093
Carac. communes						
AMII	0,577	0,494	0,000	1,000	1,000	6473093
RIP	0,341	0,474	0,000	0,000	1,000	6473093
Commune rurale	0,187	0,390	0,000	0,000	0,000	6473093
Commune de montagne	0,072	0,258	0,000	0,000	0,000	6473093
Variations en % 2014–2019						
Δ% revenus d'activité moyens	-0,019	0,922	-0,226	0,074	0,288	3923366
Δ% salaires moyens	0,035	0,935	-0,175	0,094	0,338	3751589
Δ% BIC moyens	0,156	1,204	-0,383	0,131	0,708	151990
Δ% Retraites moyennes	0,224	0,605	-0,006	0,012	0,386	2857511
Δ% Alloc, chômeages moyennes	-0,000	1,522	-0,766	0,034	0,794	520143
Δ âge moyen	3,634	8,826	5,000	5,000	5,000	6473093

tion d'une parcelle cadastrale, ainsi que sur l'âge moyen de ses habitants adultes. Nous appliquons la méthode d'étude d'évènement échelonnée à notre panel cylindré de parcelles cadastrales issues de Fidéli, pour la France entière ainsi qu'en se concentrant spécifiquement sur les parcelles situées dans des zones RIP.

La figure 3.4 présente l'évolution annuelle autour de la connexion d'un point de mutualisation à la fibre optique (année 0) du log du nombre de foyers fiscaux présents dans cette parcelle. On observe, sans restriction géographique, une tendance relativement plate avant l'année de connexion au THD, et une augmentation à partir de cette date du nombre de foyers fiscaux présents dans la parcelle, qui se stabilise rapidement et représente une augmentation de 0,1 % du nombre de foyers due à la connexion. Lorsque l'on se concentre sur les RIP, cet effet est également présent mais est plus graduel et culmine à un impact équivalent à une augmentation de 0,3 % du fait de la connexion. Dans les deux échantillons, les coefficients pré-traitement sont proches de 0, ce qui implique que les parcelles cadastrales suivent une trajectoire similaire avant de recevoir le traitement, et rend donc crédible une interprétation causale de ces effets.

Au-delà du nombre d'habitants par parcelle, on peut s'intéresser à la composition de leur population, et en particulier à l'évolution de l'âge moyen. L'usage du numérique étant négativement corrélé à l'âge, on peut en effet s'attendre à un rajeunissement de la population d'une parcelle cadastrale conséquemment à sa connexion au THD. La figure 3.5 présente les coefficients annuels de régression par étude d'évènement échelonnée sur l'âge moyen des adultes résidant au sein d'une parcelle. On observe une tendance pré-traitement plutôt plate, et un infléchissement marqué à partir de la connexion, qui est très fort dans l'échantillon complet, avec une baisse de l'âge moyen d'environ un quart d'année 3 ans après le branchement de la parcelle au THD. L'effet au sein des RIP est également négatif et significatif, mais d'une ampleur moindre, l'impact de la connexion à horizon 3 ans sur l'âge moyen correspondant plutôt à 0,1 année.

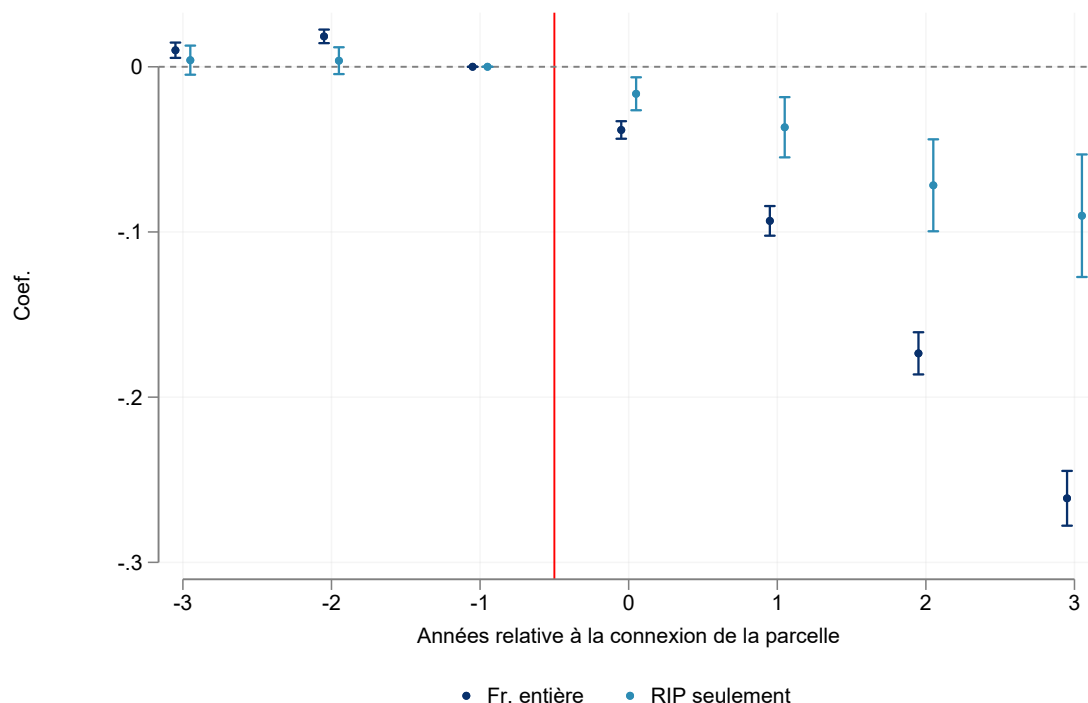
FIGURE 3.4 – Impact de la connexion THD sur le nombre de foyers fiscaux dans une parcelle cadastrale, étude d'évènement échelonnée



Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans Fidéli sur la période 2014–2019. L'estimation est réalisée par étude d'évènement échelonnée. La variable dépendante est le log du nombre de foyers fiscaux présents dans la parcelle. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale. L'estimation est réalisée sur 36,45M d'observations dans l'échantillon complet, 11,68M dans l'échantillon RIP.

Impact de la connexion au THD sur les revenus d'activité La figure 3.6 présente les résultats en étude d'évènement, utilisant comme variable dépendante le log des revenus d'activités moyens. Nous définissons comme revenus d'activité les revenus issus de salaires, ainsi que des bénéfiques professionnels (BIC, BNC, BA). On observe un profil très dynamique des revenus d'activité, dont l'infléchissement est néanmoins antérieur au branchement du PM à la fibre. Ceci suggère que le traitement que nous utilisons n'est pas réellement exogène dans l'échantillon que nous utilisons, et que les parcelles ont plus de chances d'être connectées lorsque leurs habitants connaissent une phase de croissance des revenus (par exemple parce que

FIGURE 3.5 – Impact de la connexion THD sur l'âge moyen des adultes dans une parcelle cadastrale, étude d'évènement échelonnée



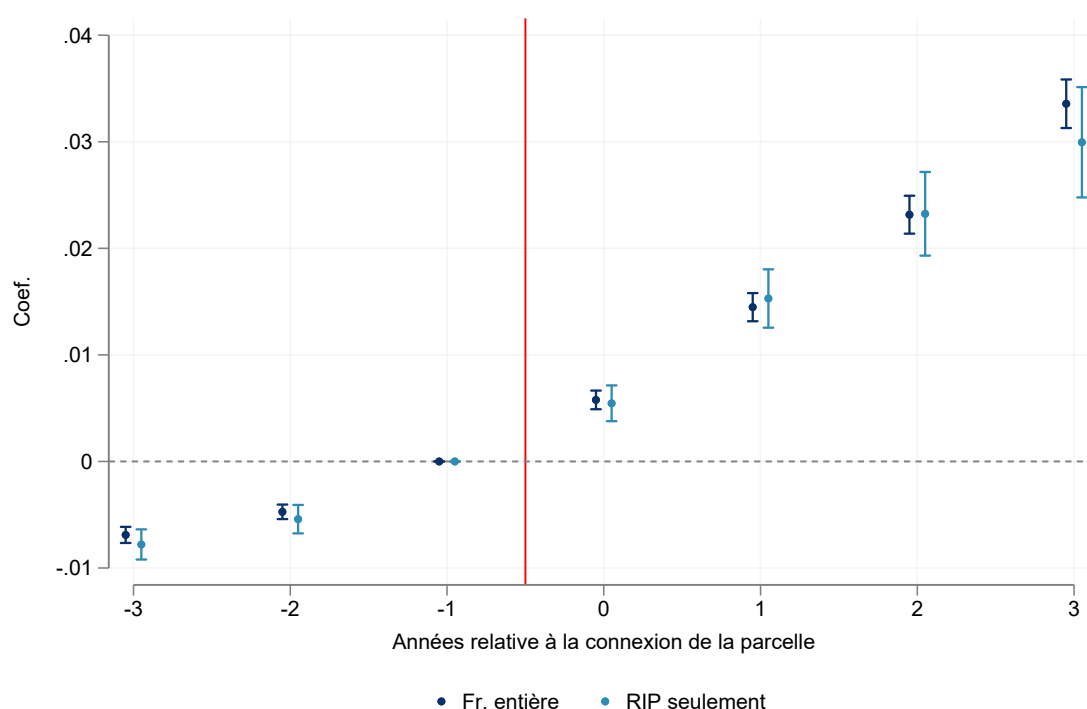
Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans Fidéli sur la période 2014-2019. L'estimation est réalisée par étude d'évènement échelonnée. La variable dépendante est l'âge moyen des adultes (redevables fiscaux) présents dans la parcelle. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale. L'estimation est réalisée sur 36,50M d'observations dans l'échantillon complet, 11,69M dans l'échantillon RIP.

ces parcelles connaissent un rajeunissement de leur population antérieurement à la connexion fibre). Ces résultats ne sont donc pas interprétables comme un impact causal de la connexion à la fibre sur les revenus d'activité.

Les régressions par étude d'évènement étant peu satisfaisantes, nous menons une étude d'évènement empilée avec sélection explicite du groupe de contrôle. Si cette méthode a le défaut de réduire considérablement la taille d'échantillon (seules les parcelles pour lesquelles une parcelle contrôle suffisamment proche a pu être trouvée), elle permet néanmoins de rendre a priori plus comparables les groupes de contrôle et de traitement. La figure 3.7 présente ainsi des résultats similaires en

utilisant la méthode de régression empilée. On observe cette fois une tendance plutôt décroissante des revenus sur l'ensemble de la période pour l'échantillon France entière, et une tendance plutôt plate dans les RIP. Au global, l'analyse de ces résultats ne permet pas de mettre en évidence un effet de la connexion au THD sur les revenus d'activité moyens au niveau des parcelles cadastrales.

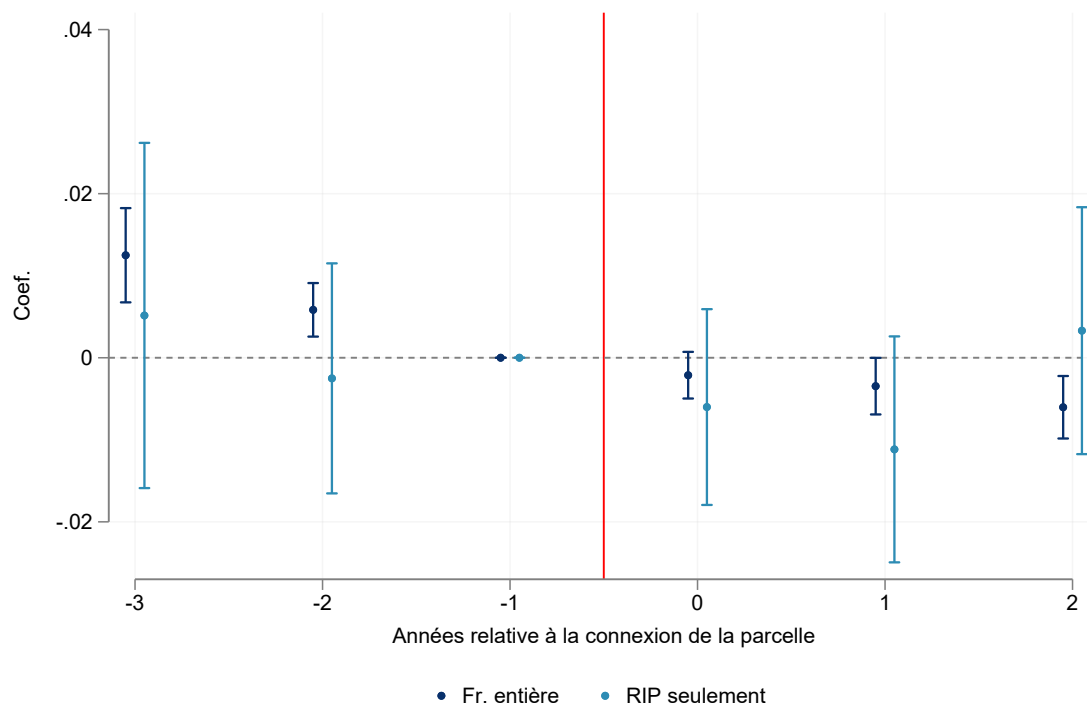
FIGURE 3.6 – Impact de la connexion THD sur les revenus d'activité moyens (log) dans une parcelle cadastrale, étude d'évènement échelonnée



Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans Fidéli sur la période 2014–2019. L'estimation est réalisée par étude d'évènement échelonnée. La variable dépendante est le log de la moyenne des revenus d'activité déclarés par les foyers fiscaux dans la parcelle. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale. L'estimation est réalisée sur 24,39M d'observations dans l'échantillon complet, 7,69M dans l'échantillon RIP.

Une hypothèse plus précise sur les effets potentiels de la connexion à la fibre est que celle-ci a pu faciliter l'exercice de professions d'indépendants dont les besoins en bande passante sont importants (développeurs, designers, photographes, etc.). On peut donc étudier l'évolution autour de la connexion à la fibre de la probabilité qu'un foyer déclare des revenus industriels et commerciaux, qui incluent les

FIGURE 3.7 – Impact de la connexion THD sur les revenus d'activité moyens (log) dans une parcelle cadastrale, étude d'évènement empilée



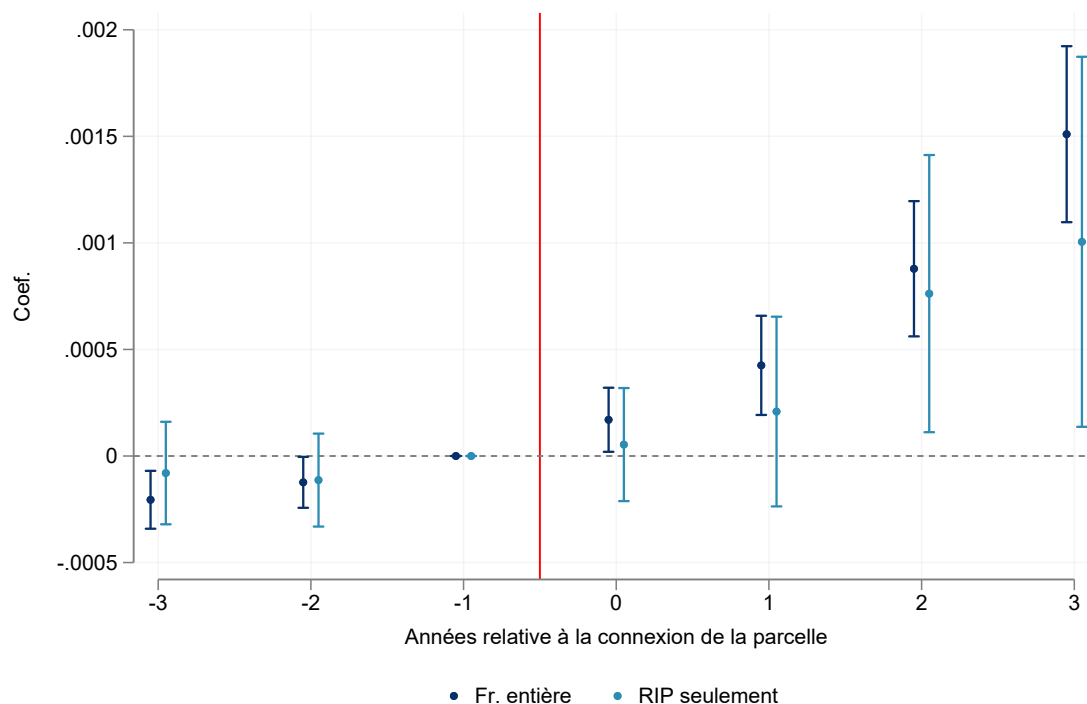
Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans Fidéli sur la période 2014–2019. L'estimation est réalisée par étude d'évènement empilée avec sélection explicite du groupe de contrôle. La variable dépendante est le log de la moyenne des revenus d'activité déclarés par les foyers fiscaux dans la parcelle. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale. L'estimation est réalisée sur 5,54M d'observations dans l'échantillon complet, 461k dans l'échantillon RIP.

activités dont on peut faire l'hypothèse qu'elles soient sensibles à l'existence d'une connexion internet de qualité.

La figure 3.8 présente l'évolution de cette probabilité autour de la connexion THD, estimée par étude d'évènement. On observe cette fois des tendances plutôt plates avant la connexion de la parcelle, tant dans l'échantillon complet que dans l'échantillon RIP. On constate ensuite une augmentation graduelle dans les deux échantillons, qui culmine autour de + 0,1 pp dans les RIP et + 0,15 pp en France entière de chances qu'un foyer ait des revenus industriels et commerciaux. Ceci se compare à une moyenne de 4,5 % de cette probabilité l'année précédant la connex-

tion (tableau 3.2), ce qui signifie que l'augmentation est représentée entre 2 et 3 % de la moyenne initiale.

FIGURE 3.8 – Impact de la connexion THD sur la probabilité d'avoir des revenus industriels et commerciaux dans une parcelle cadastrale, étude d'évènement échelonnée



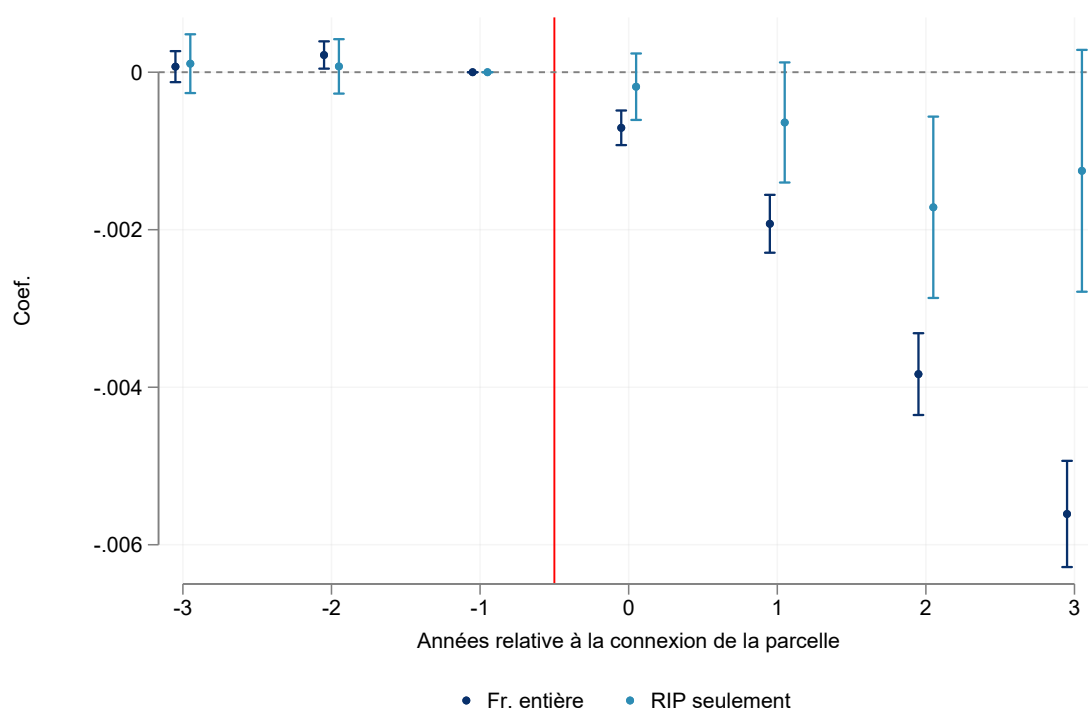
Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans Fidéli sur la période 2014-2019. L'estimation est réalisée par étude d'évènement échelonnée. La variable dépendante est la probabilité que des bénéfices industriels et commerciaux soient déclarés par au moins un foyer fiscal dans la parcelle. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale. L'estimation est réalisée sur 36,50M d'observations dans l'échantillon complet, 11,69M dans l'échantillon RIP.

Impact de la connexion au THD sur les revenus de remplacement Une dernière dimension selon laquelle la connexion d'une parcelle au THD est susceptible d'affecter la structure des revenus est la perception de revenus de remplacement, comme les pensions de retraite et les allocations chômage.

La figure 3.9 présente les coefficients de régression en utilisant comme variable dépendante la probabilité de percevoir des pensions de retraite. En cohérence avec les résultats sur l'âge moyen présentés en figure 3.5. Cet effet est plus fort sur

l'échantillon France entière, pour lequel la connexion représente une baisse de 0,6 pp de percevoir des pensions, contre environ une baisse de 0,2 pp pour l'échantillon RIP, soit des baisses respectives d'environ 1 % et 0,4 %.

FIGURE 3.9 – Impact de la connexion THD sur la probabilité de percevoir des pensions de retraite dans une parcelle cadastrale, étude d'évènement échelonnée

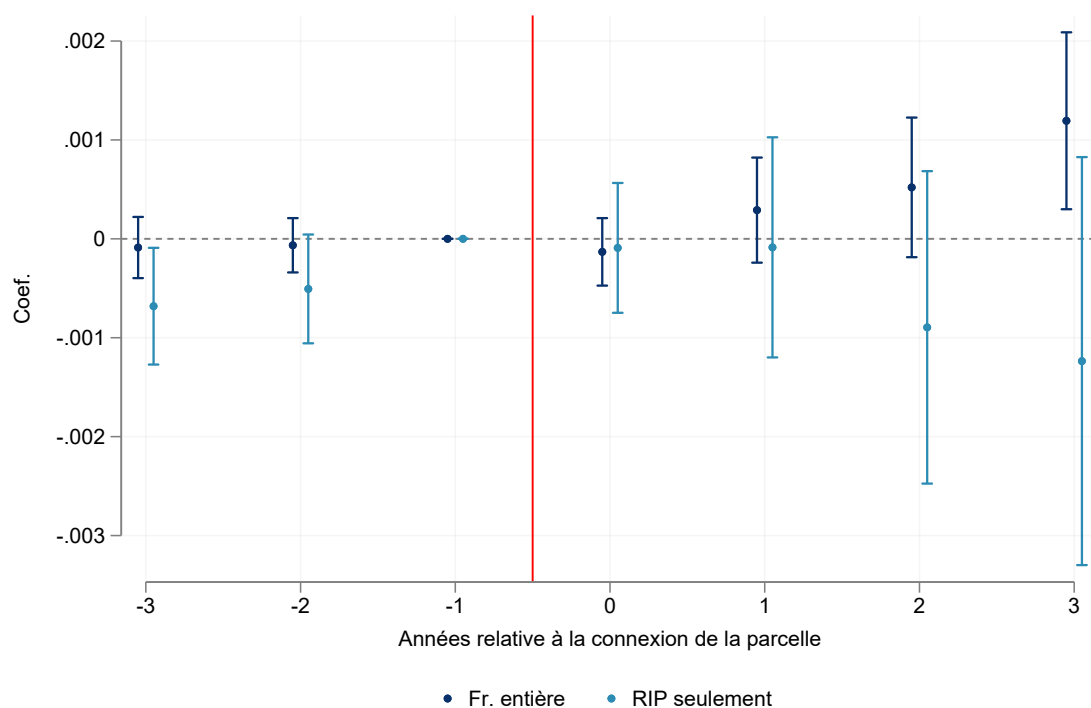


Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans Fidéli sur la période 2014–2019. L'estimation est réalisée par étude d'évènement échelonnée. La variable dépendante est la probabilité que des pensions de retraites soient perçues par au moins un foyer dans la parcelle. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale. L'estimation est réalisée sur 24,39M d'observations dans l'échantillon complet, 7,69M dans l'échantillon RIP.

De manière similaire, la figure 3.10 estime l'effet de la connexion fibre sur la probabilité qu'un foyer dans la parcelle déclare des allocations chômage. On peut faire l'hypothèse que deux effets peuvent être produits par la disponibilité de la fibre : d'une part, une connexion internet de qualité peut faciliter la recherche d'emploi ; d'autre part, celle-ci peut également faciliter le recours aux allocations chômage, allant à l'encontre du premier effet. Empiriquement, nous ne mettons

néanmoins en évidence aucun effet clair : l'échantillon France entière témoigne d'un effet légèrement positif mais peu significatif statistiquement. L'échantillon RIP présente à l'inverse un effet négatif mais également non significatif statistiquement. On peut donc exclure que la connexion au THD ait un effet de premier ordre sur la probabilité de bénéficier d'allocations chômage.

FIGURE 3.10 – Impact de la connexion THD sur la probabilité de bénéficier d'allocations chômage dans une parcelle cadastrale, étude d'évènement échelonnée



Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans Fidéli sur la période 2014–2019. L'estimation est réalisée par étude d'évènement échelonnée. La variable dépendante est la probabilité que des allocations chômage soient perçues par au moins un foyer dans la parcelle. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale. L'estimation est réalisée sur 24,39M d'observations dans l'échantillon complet, 7,69M dans l'échantillon RIP.

3.4 Impact de la connexion au THD sur le marché immobilier local

Dans cette section, nous examinons au niveau local différentes variables d'intérêt ayant trait à la dynamique du marché immobilier. Comme dans la section précédente, nous adoptons principalement une approche par étude d'évènements échelonnée qui est parfois complétée par une approche par étude d'évènement empilée (voir section 3.2.1 et 3.2.2 pour une description générale de ces deux approches).

Les statistiques descriptives concernant l'échantillon complet sont présentées dans le tableau 3.3. Le panneau supérieur présente les caractéristiques des parcelles un an avant le déploiement de la fibre. Le taux de vacance moyen des logements est de 5 %. Il s'agit d'une variable continue mais dont, en pratique, la distribution au niveau parcelle a une masse largement concentrée sur les valeurs 0 et 1⁷. Le même constat s'applique au taux en location qui décrit la part des logements qui sont en location au sein d'une parcelle donnée. Environ 6 % des parcelles abritent un établissement avec un code NAF valide (Étab avec NAF (indic.)). Nous utilisons cette variable afin de capter la présence et l'implantation de nouveaux établissements au niveau local.

Une hausse de l'attractivité d'une zone donnée suite à la connexion à la fibre devrait se manifester par une hausse de la demande pour les biens immobiliers qui se trouvent dans cette zone. Cette hausse de la demande peut mener à un ajustement par les quantités ainsi que par les prix—l'importance relative de ces ajustements est susceptible de varier selon l'élasticité de l'offre locale (Glaeser et Gyourko, 2018). Nous considérons en premier lieu des variables captant les ajustement par les quantités avant d'examiner les dynamiques des prix.

7. Nous nous concentrons sur la vacance des logements et non pas des locaux d'activité pour des questions de qualité de données (Cerema, 2020)

TABLEAU 3.3 – Statistiques descriptives au niveau parcelle cadastrale, échantillon complet

	(1)					
	mean	sd	p25	p50	p75	N
Carac. parcelles en t-1						
Taux de vacance (logement)	0,058	0,217	0,000	0,000	0,000	13713463
Taux en location	0,186	0,366	0,000	0,000	0,000	15527028
# locaux d'activité	0,183	2,695	0,000	0,000	0,000	15527028
# maisons	0,930	1,516	1,000	1,000	1,000	15527028
# appartements	1,092	8,909	0,000	0,000	0,000	15527028
Taux occupé par propriétaire	0,795	0,385	1,000	1,000	1,000	14873362
Étab, avec NAF (indic,)	0,067	0,249	0,000	0,000	0,000	15527028
Loué (indic,)	0,229	0,420	0,000	0,000	0,000	15527028
Mutations DV3F (indic,)	0,055	0,228	0,000	0,000	0,000	15527028
Carac. communes						
AMII	0,616	0,486	0,000	1,000	1,000	15523902
RIP	0,272	0,445	0,000	0,000	1,000	15523902
Commune rurale	0,144	0,351	0,000	0,000	0,000	15523902
Commune de montagne	0,065	0,247	0,000	0,000	0,000	15523902
Variations en % ou niveau 2014–2019						
Δ% Étab, avec NAF (indic,)	-0,019	0,317	0,000	0,000	0,000	866604
Δ% locations	-0,007	0,198	0,000	0,000	0,000	3419022
Δ% mutations	0,063	0,862	-0,405	0,000	0,606	177144
Δ taux en location	-0,004	0,110	0,000	0,000	0,000	15527028
Δ taux de vacances	0,007	0,258	0,000	0,000	0,000	8194722
Observations	15527028					

Notes : Ce tableau présente des statistiques descriptives concernant l'échantillon d'estimation principale de l'approche d'étude d'évènements échelonnée sur le marché immobilier. Le taux de vacance des logements est défini comme le ratio du nombre de logements vacants sur le nombre de logements total. Le taux en location est la ratio du nombre de logement offert en location sur le nombre total de logement. Le terme # indique le dénombrement de la variable correspondante au niveau de la parcelle cadastrale. Le terme "Indic." indique que la variable décrite est une indicatrice.

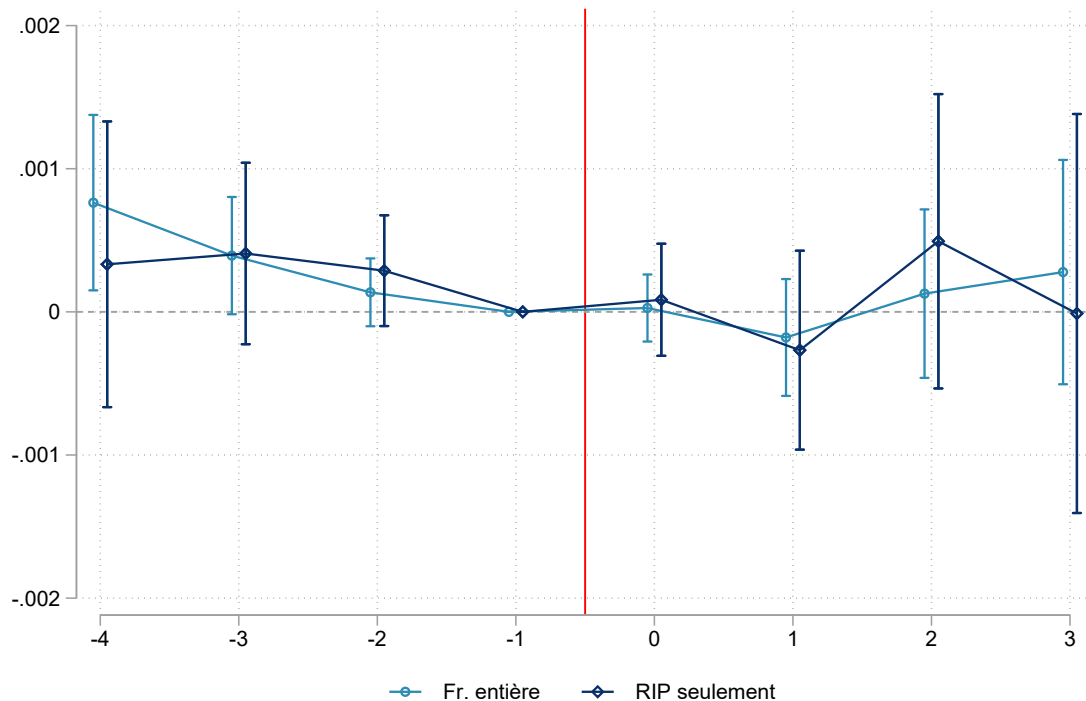
Impact de la connexion au THD sur le taux de vacance des logements et l'occupation par des entreprises Dans un premier temps, nous analysons l'impact de la connexion au THD sur le taux de vacance des logements, sur la part des logements en location ainsi que sur la présence d'entreprises. Comme ci-dessus, nous appliquons la méthode d'étude d'évènements échelonnée à notre panel cylindré de parcelles cadastrales issues des Fichiers fonciers, pour la France entière ainsi qu'en se concentrant spécifiquement sur les parcelles situées dans des zones RIP.

La figure 3.11 présente l'évolution annuelle autour de la connexion d'un point de mutualisation à la fibre optique (année 0) du taux de vacance des logements des parcelles. On observe, sans restriction géographique, une légère tendance à la baisse avant l'année de connexion au THD, et un aplatissage de la tendance, et les coefficients cessent d'être différents de zéro. Le même constat s'applique à l'échantillon des parcelles situées en communes RIP. Dans les deux échantillons, on ne détecte pas de différence significative en termes de taux de vacance. Nous pouvons, de plus, rejeter des effets de taille modeste : il est possible de rejeter un effet de -0,05 point de pourcentage, soit -0,8 % par rapport à la moyenne de l'échantillon, avec un niveau de confiance de 95% dans le cas de l'échantillon total.

On considère maintenant l'approche empilée sur l'échantillon complet (la part des RIP est très faible dans la base empilée d'où notre focus sur l'échantillon complet). La figure 3.12 suggère un effet négatif très limité de -0,2 point de pourcentage. Néanmoins il est délicat d'interpréter ce coefficient de manière causale, dans la mesure où les coefficients avant le déploiement de la fibre sont estimés de façon très bruitée, de telle sorte que le fait que l'on ne soit pas en mesure de rejeter des pré-tendances nulles est peu informatif (du fait du manque de puissance statistique).

On considère maintenant la part des logements qui sont proposés à la location (et non pas occupés par leurs propriétaires), indépendamment du fait qu'ils soient occupés ou vacants. On remarque un effet négatif d'environ -0,05 point de

FIGURE 3.11 – Impact de la connexion THD sur le taux de vacance des logements (parcelle cadastrale), étude d'évènement échelonnée

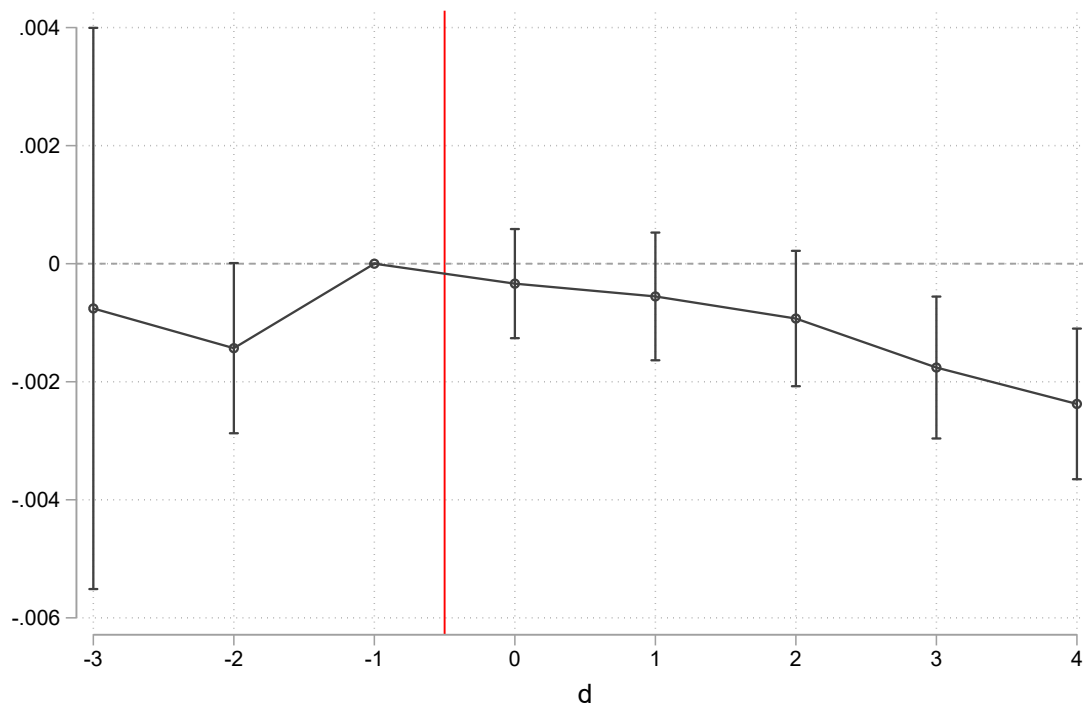


Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans les Fichiers fonciers sur la période 2013–2019. L'estimation est réalisée par étude d'évènement échelonnée. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale (voir équation (3.1) pour plus de détails). Les intervalles de confiance à 95 % sont représentés autour des coefficients. Les erreurs-standard sont groupées au niveau parcelle cadastrale.. La variable dépendante est le taux de vacance des logements. L'estimation est réalisée sur 49,61M d'observations dans l'échantillon complet, 16,78M dans l'échantillon RIP.

pourcentage trois ans après l'année de la connexion. Les coefficients pré-connexion étant relativement proches de 0, une interprétation causale semble plausible. Néanmoins, cet effet, bien que statistiquement significatif, est très faible d'un point de vue économique puisque qu'environ 19 % des logements sont proposés à la location dans l'échantillon en moyenne l'année avant la connexion, de telle sorte que l'effet représente environ 0,26 % de la moyenne pré-traitement.

La figure 3.14 présente les résultats analogues pour la spécification empilée avec des résultats très proches, avec un coefficient d'environ -0,0005 deux ans après le déploiement.

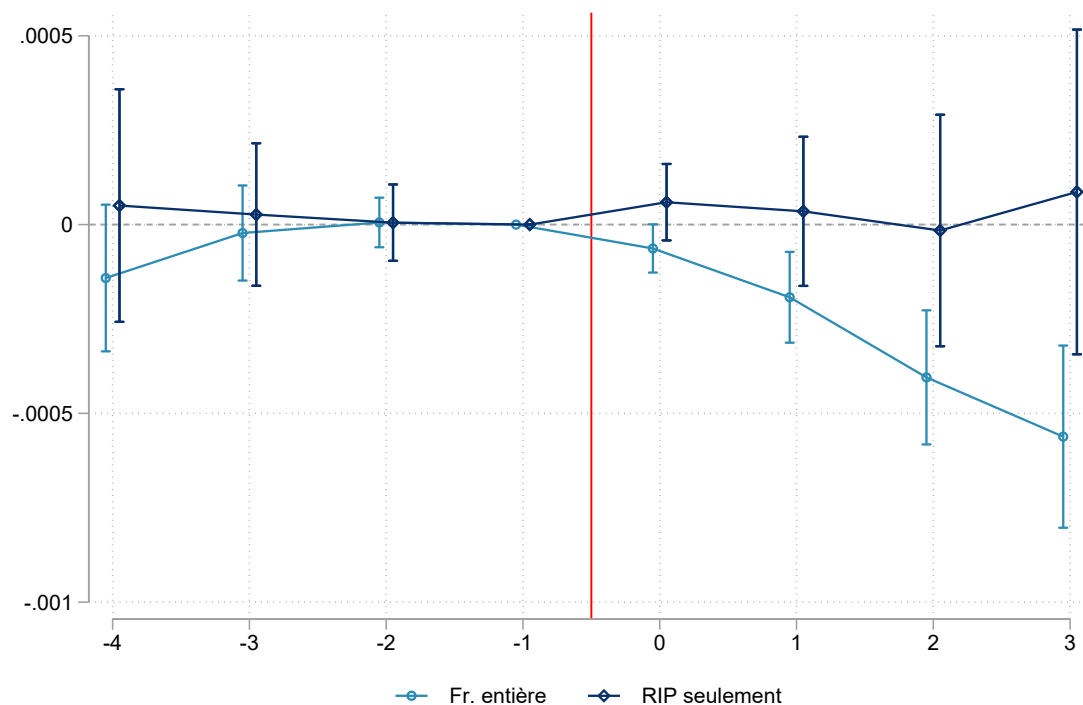
FIGURE 3.12 – Impact de la connexion THD sur le taux de vacance des logements (parcelle cadastrale), étude d'évènement empilée



Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans les Fichiers fonciers sur la période 2013–2019. L'estimation est réalisée par étude d'évènement échelonnée. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale (voir équation (3.3) pour plus de détails). Les intervalles de confiance à 95 % sont représentés autour des coefficients. Les erreurs-standard sont groupées au niveau parcelle cadastrale. La variable dépendante est le taux de vacance des logements. L'estimation est réalisée sur 26,17M d'observations.

Nous nous intéressons maintenant à l'attractivité du point de vue des entreprises en considérant la probabilité qu'une parcelle abrite une entreprise. La Figure 3.15 montre les résultats de l'approche échelonnée. Concernant l'échantillon entier, les coefficients pré-traitement suggèrent une violation assez nette de l'hypothèse de tendances communes puisque le coefficient est significativement négatif pour les périodes $d = -4$ à $d = -2$. On constate la poursuite de ces dynamiques divergentes post-traitement sans que l'on ne puisse déceler de rupture par rapport à la tendance pré-traitement. Ces résultats suggèrent donc que le déploiement de la fibre a eu lieu prioritairement parmi des parcelles qui exhibaient un plus fort dynamisme

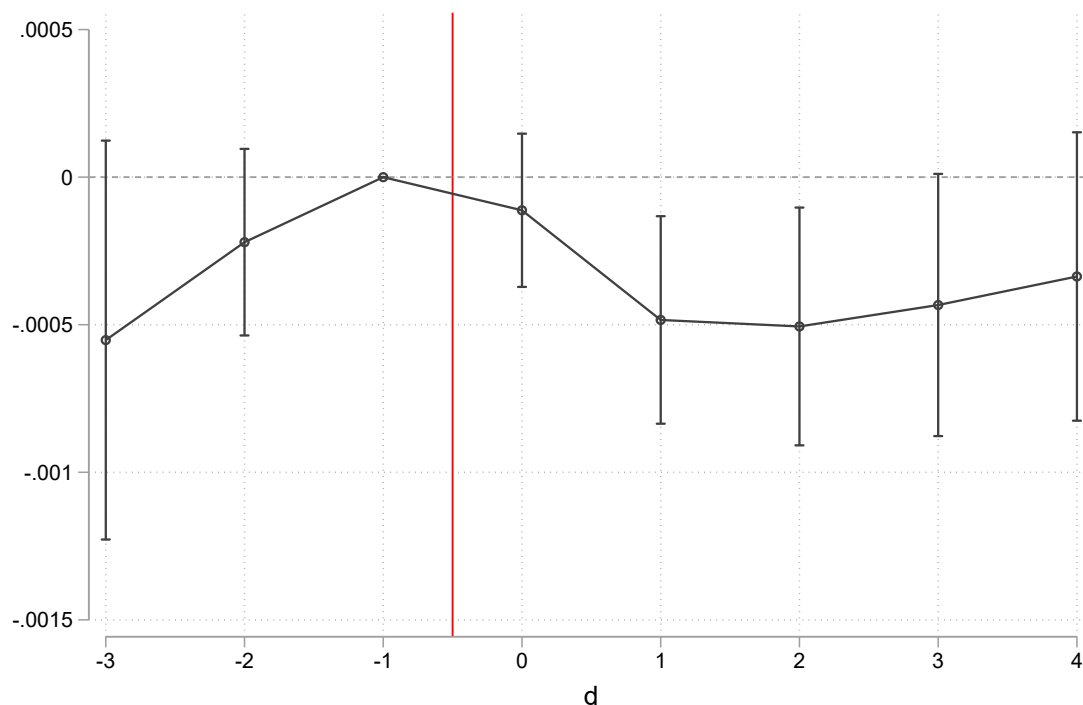
FIGURE 3.13 – Impact de la connexion THD sur la part de logement en location, étude d'évènement échelonnée



Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans les Fichiers fonciers sur la période 2013–2019. L'estimation est réalisée par étude d'évènement échelonnée. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale (voir équation (3.1) pour plus de détails). Les intervalles de confiance à 95 % sont représentés autour des coefficients. Les erreurs-standard sont groupées au niveau parcelle cadastrale. La variable dépendante est la part de logement en location. L'estimation est réalisée sur 49,61M d'observations dans l'échantillon complet, 16,78M dans l'échantillon RIP.

en termes d'implantations d'entreprises, et vient dès lors poser question sur l'exogénéité du traitement dans ce contexte. On peut noter cependant que ces dynamiques divergentes, bien que précisément documentées d'un point de vue statistique, sont d'une importance quantitative limitée. En effet, la déviation de 0,00075 indiqué par le coefficient à $d = -4$ ne correspondant qu'à une déviation équivalent à 1,11 % de la moyenne de l'échantillon (de 6,7 %). Dans le cas des parcelles situées au sein de communes RIP, les coefficients suggèrent une sélection sur la dynamique moindre que dans l'échantillon totale. On ne détecte cependant pas d'effet notable dans les périodes très traitement.

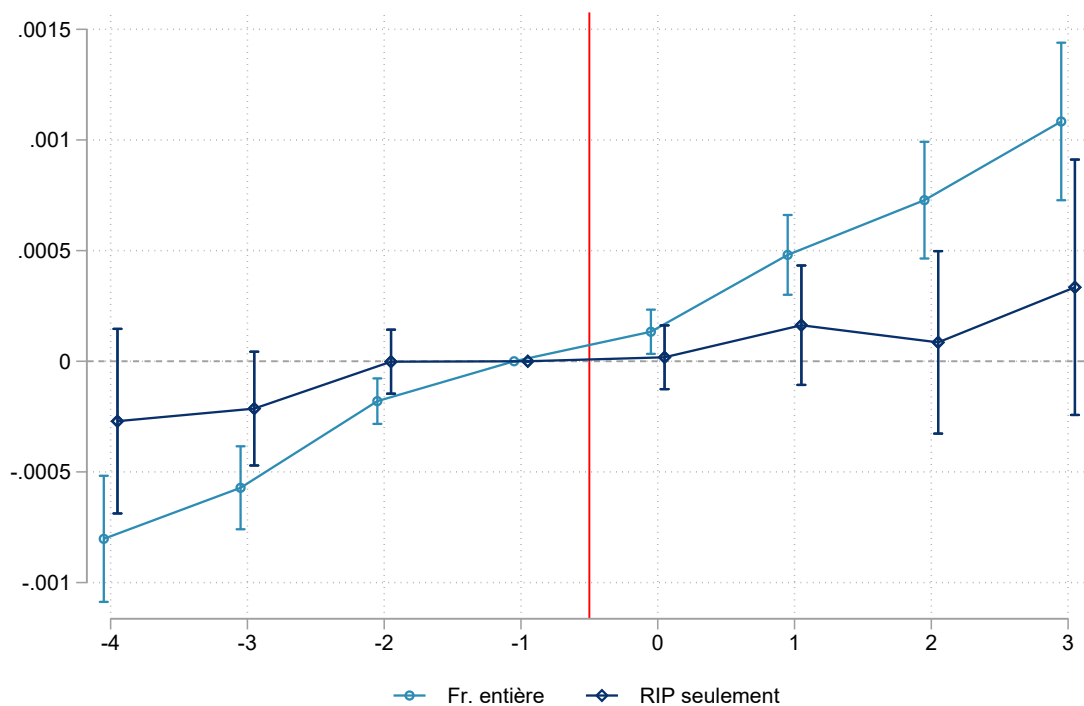
FIGURE 3.14 – Impact de la connexion THD sur le taux de logement en location, étude d'évènement empilée



Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans les Fichiers fonciers sur la période 2013–2019. L'estimation est réalisée par étude d'évènement échelonnée. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale (voir équation (3.3) pour plus de détails). Les intervalles de confiance à 95 % sont représentés autour des coefficients. Les erreurs-standard sont groupées au niveau parcelle cadastrale.. La variable dépendante est la part de logement en location. L'estimation est réalisée sur 26,170M d'observations.

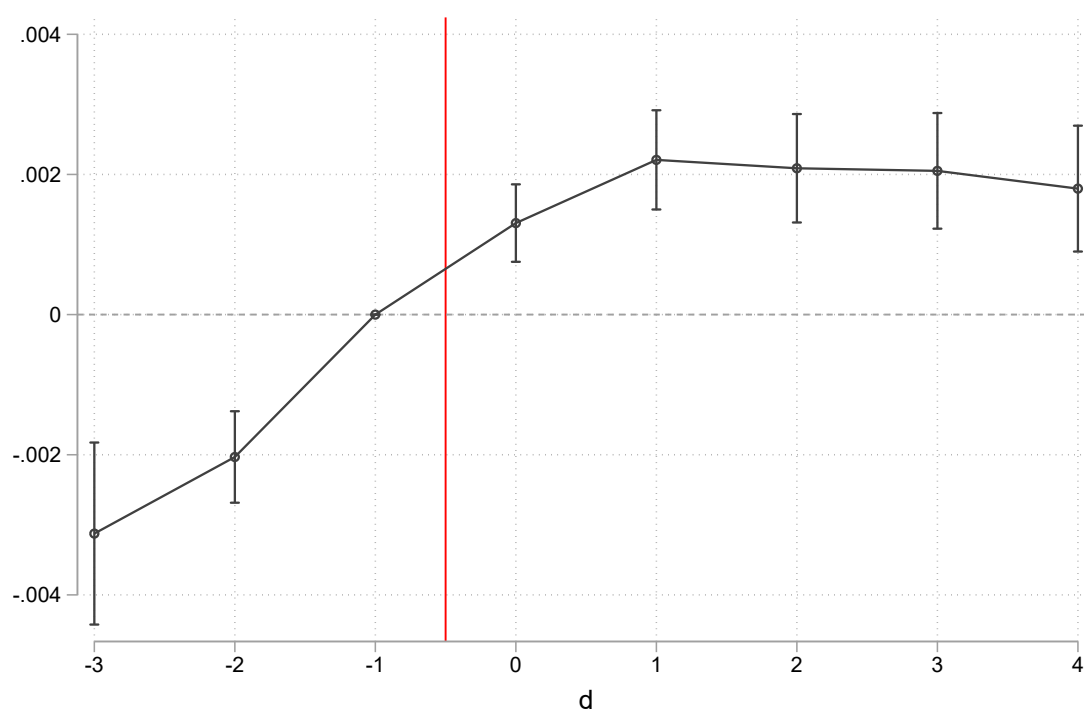
La figure 3.16 présente les résultats analogues pour la spécification empilée avec des résultats qualitativement similaire, avec un coefficient d'environ -0,002 deux ans avant le déploiement et +0,002 deux après le traitement. L'approche empilée ne semble donc pas, dans ce cas précis, à même de mieux neutraliser des différences de tendances entre groupes traité et contrôle.

FIGURE 3.15 – Impact de la connexion THD sur la probabilité de présence d’une entreprise, étude d’évènement échelonnée



Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans les Fichiers fonciers sur la période 2013–2019. L’estimation est réalisée par étude d’évènement échelonnée. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale (voir équation (3.1) pour plus de détails). Les intervalles de confiance à 95 % sont représentés autour des coefficients. Les erreurs-standard sont groupées au niveau parcelle cadastrale. La variable dépendante est une indicatrice égale à 1 lorsque la parcelle contient au moins une entreprise dont le code NAF est correctement renseigné au 1er janvier de l’année en question. L’estimation est réalisée sur 49,61M d’observations dans l’échantillon complet, 16,78M dans l’échantillon RIP.

FIGURE 3.16 – Impact de la connexion THD sur la probabilité de présence d'une entreprise, étude d'évènement empilée



Notes : Cette figure présente les coefficients de régression estimés par année relative au traitement, sur un panel cylindré de parcelles cadastrales présentes dans les Fichiers fonciers sur la période 2013–2019. L'estimation est réalisée par étude d'évènement échelonnée. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". Les erreurs-standard sont groupées au niveau parcelle cadastrale (voir équation (3.3) pour plus de détails). Les intervalles de confiance à 95 % sont représentés autour des coefficients. Les erreurs-standard sont groupées au niveau parcelle cadastrale. La variable dépendante est une indicatrice égale à 1 lorsque la parcelle contient au moins une entreprise dont le code NAF est correctement renseigné au 1er janvier de l'année en question. L'estimation est réalisée sur 26.170M d'observations.

Impact de la connexion au THD sur la valeur des transactions de logements

Dans cette partie, nous ré-examinons l'effet de la connexion à la fibre sur la valeur des logements connectés. À cette fin, nous estimons un modèle analogue d'étude d'évènement échelonnée et empilée aux sections précédentes à l'exception près que le niveau d'observation n'est pas ici la parcelle cadastrale mais la transaction (mutation). Cette différence est mineure puisque le statut de traitement (accès à la fibre) est défini au niveau de la parcelle cadastrale. Néanmoins, travailler au niveau des transactions présente l'avantage de pouvoir intégrer des contrôles captant les caractéristiques des biens immobiliers faisant l'objet d'une transaction (nombre d'étages, nombre de salles de bain etc.). Notre analyse se concentre sur les logements dont la valeur et la surface sont bien définis. Nous nous concentrons également sur les mutations liées à des ventes.

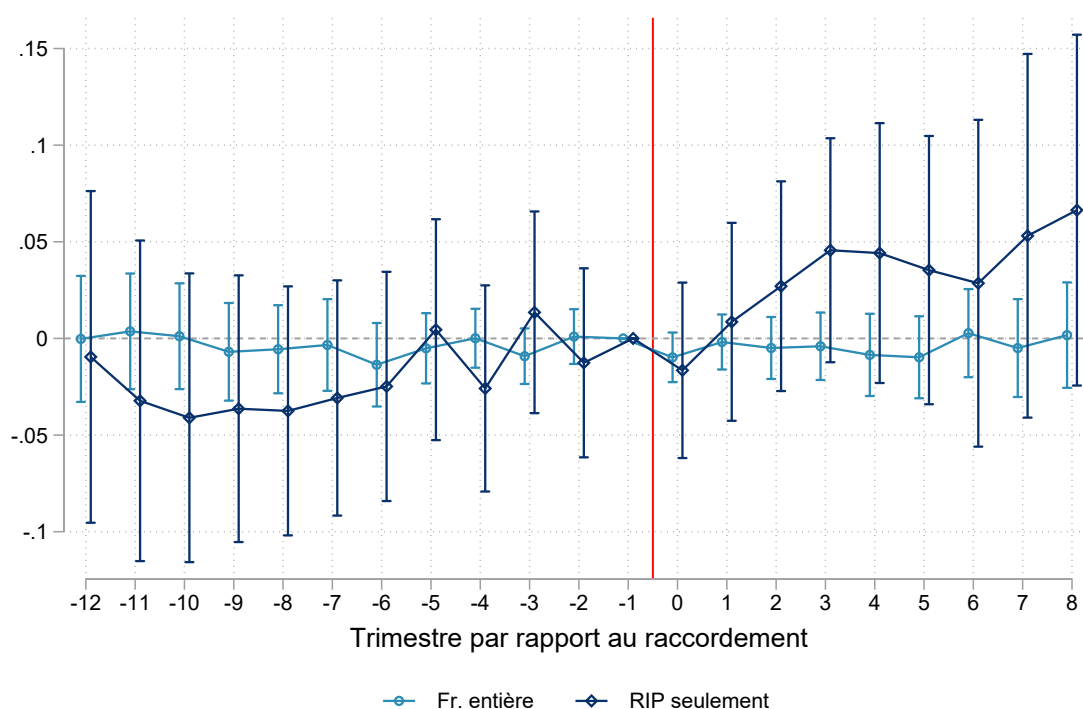
Les données Arcep sur le déploiement de la fibre nous permettent d'identifier la date exacte à laquelle une parcelle a été connectée (selon la définition présentée dans la section 3.1). Les données DV3F documentent également la date précise de la mutation. Nous sommes donc en mesure d'adopter une approche à une fréquence plus élevée que la fréquence annuelle. Nous choisissons la fréquence trimestrielle qui permet de prendre en compte l'aspect potentiellement saisonnier des prix immobiliers (voir par exemple Rosenthal, 2006).

La figure 3.17 montre les résultats de l'estimation lorsque l'on considère le logarithme de la valeur totale comme variable dépendante. On constate pour l'échantillon totale une forte stabilité des coefficients à la fois avant et après le traitement, ce qui soutient à la fois la comparabilité des dynamiques des groupes traité et contrôle ainsi que l'absence d'effet. Nos résultats sont relativement précis dans la mesure où l'intervalle de confiance à 95 % exclut des effets de l'ordre de +2,5 %. Néanmoins, cet intervalle n'exclut pas des effets de l'ordre de +1 % tels qu'estimés par Ahlfeldt et al. (2017) au Royaume-Uni.⁸

8. L'effet de 1 % est associé au passage de 8 à 24 Mbit/s.

La figure 3.18 présente les résultats en utilisant le prix au mètre carré et trouvent des résultats qualitativement identiques à ceux de la figure 3.17. Enfin, la figure 3.19 présente les résultats sur la base d'un prix au carré résidualisé vis-à-vis de caractéristiques privées du logement⁹.

FIGURE 3.17 – Impact de la connexion THD sur la valeur des transactions, étude d'évènement échelonnée

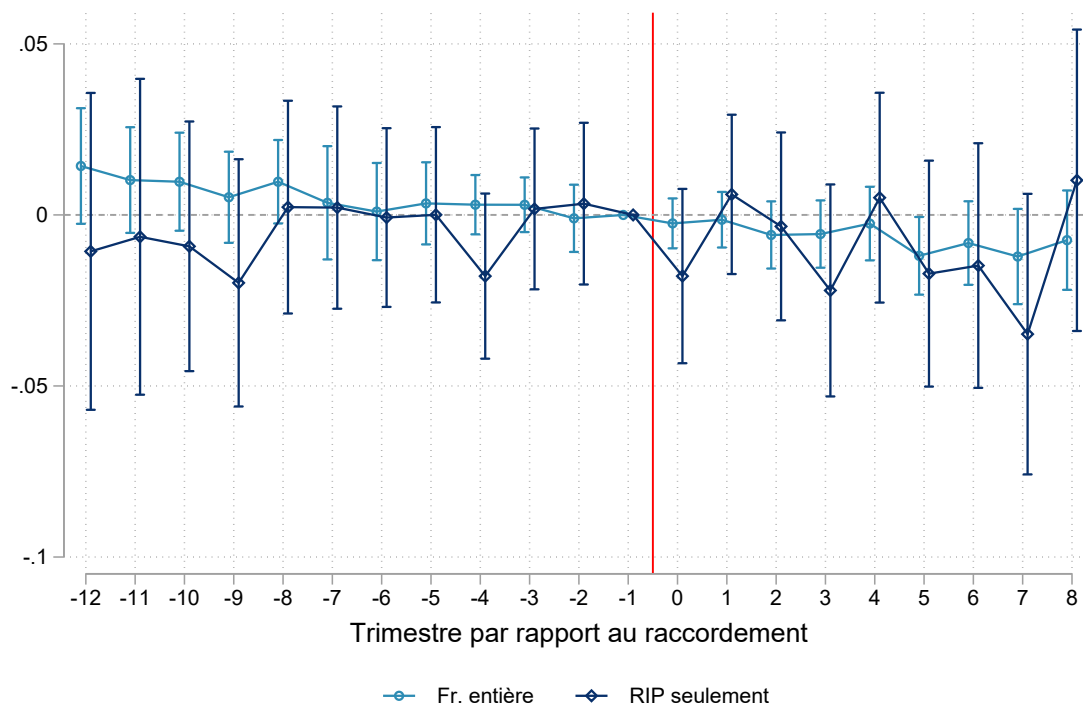


Notes : Cette figure présente les coefficients de régression estimés par trimestre relatif au traitement, sur un panel non-cylindrique de transactions de biens immobiliers (logements) sur la période 2010-2019. L'estimation est réalisée par étude d'évènement échelonnée. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". La variable dépendante est le log de la valeur foncière totale associée à la transactions. Les erreurs-standard sont groupées au niveau parcelle cadastrale. L'estimation est réalisée sur 26,17M d'observations.

Il est possible que les différences entre parcelles soient perçues comme temporaires, ce qui pourrait expliquer l'absence d'effet clair capté par notre approche empirique locale sur les prix. Cela permettrait d'expliquer pourquoi nos résultats diffèrent de Ahlfeldt et al. (2017) dans la mesure où leur stratégie empirique exploite des différences locales mais permanentes de débit internet.

9. La résidu est obtenu sur la base d'une régression hédonique (Poulhès, 2015) qui inclut les caractéristiques suivantes : log de la surface totale, part de la surface en appartement, nombre d'étages (indicatrices) et nombre de pièces (indicatrices).

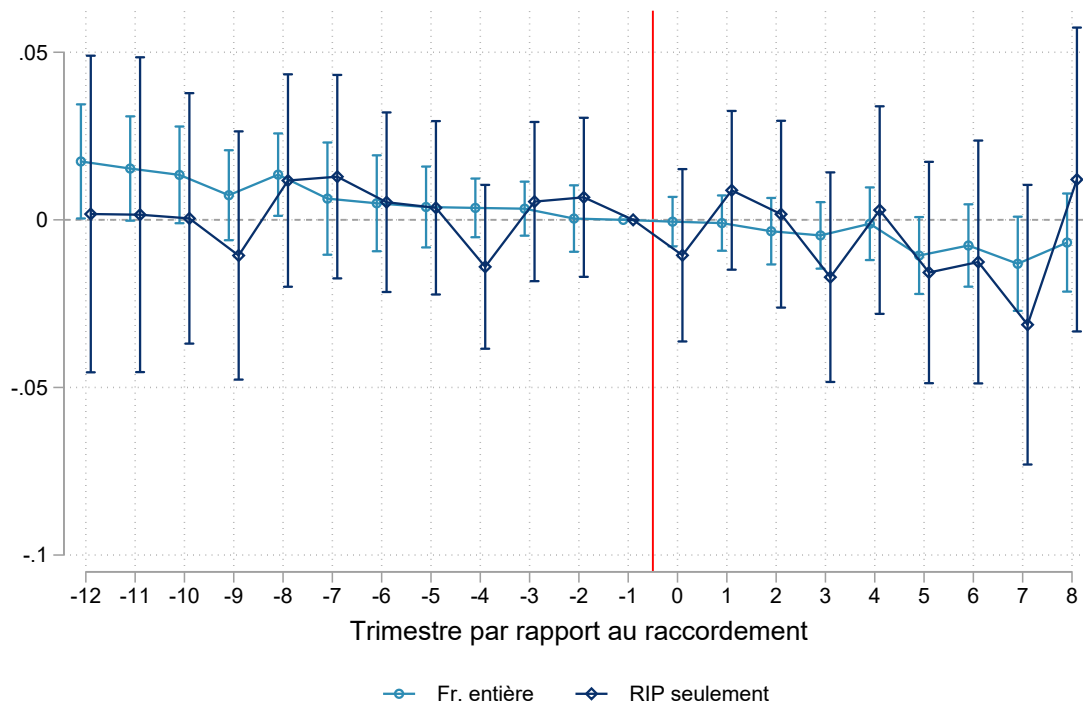
FIGURE 3.18 – Impact de la connexion THD sur la valeur au mètre carré des transactions, étude d'évènement échelonnée



Notes : Cette figure présente les coefficients de régression estimés par trimestre relatif au traitement, sur un panel non-cylindrique de transactions de biens immobiliers (logements) sur la période 2010-2019. L'estimation est réalisée par étude d'évènement échelonnée. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". La variable dépendante est le log de la valeur foncière par mètre carré (prix ajusté pour la surface) associée à la transactions. Les erreurs-standard sont groupées au niveau parcelle cadastrale. L'estimation est réalisée sur 26,17M d'observations.

Une autre différence majeure avec entre notre étude et Ahlfeldt et al. (2017) est le type de raccordements étudiés. Ahlfeldt et al. (2017) étudient des déploiements de l'ADSL avec d'une part des passages à 8 mbit/s et d'autres part des passages de 8 à 24 mbit/s. Ahlfeldt et al. (2017) montrent que le rendement de la vitesse est largement décroissant avec des effets de +2,5 % pour le premier type de transitions et +1 % pour le second type de transitions. Nous étudions des connexions qui permettent des transitions vers un débit plus élevé grâce à la fibre, néanmoins, il est possible que, au moins au sein des parcelles situés au sein de zones assez denses, la connexion pré-fibre soient plus élevées que dans le cas de l'étude de Ahlfeldt et al. (2017). Dès lors, cette absence d'effet net pourrait être expliquée par

FIGURE 3.19 – Impact de la connexion THD sur la valeur au mètre carré résidualisée des transactions, étude d'évènement échelonnée



Notes : Cette figure présente les coefficients de régression estimés par trimestre relatif au traitement, sur un panel non-cylindrique transactions de biens immobiliers (logements) sur la période 2010–2019. L'estimation est réalisée par étude d'évènement échelonnée. La régression inclut des effets fixes "département × année" ainsi que des effets-fixes "parcelle cadastrale". La variable dépendante est le log de la valeur par mètre carré résidualisée vis-à-vis de caractéristiques du logement associée à la transactions. Les erreurs-standard sont groupées au niveau parcelle cadastrale. L'estimation est réalisée sur 26,17M d'observations.

des rendements décroissants du débit de l'internet fixe.

3.5 Conclusion

Ce chapitre s'est attaché à analyser à un niveau spatial très fin (la parcelle cadastrale), les effets d'une connexion à la fibre (mesurée comme le raccordement d'un point de mutualisation) sur des dimensions comme la population vivant dans la parcelle, les revenus de ses habitants, et le marché immobilier local.

Nos résultats indiquent un impact positif, que l'on peut interpréter comme étant causal, de la connexion au THD, sur la démographie des parcelles cadastrales traitées, qui voient le nombre de leurs foyers fiscaux augmenter d'environ 0,1 % en général, et 0,3 % dans les RIP. Cet effet s'accompagne d'une baisse de l'âge moyen des adultes, qui est en revanche plus fort sur l'ensemble du territoire (-0,25 année) que dans les territoires RIP (-0,1 année). Le raccordement au THD a également un impact positif, quoique de faible ampleur, sur la probabilité que les foyers génèrent des revenus industriels et commerciaux, et un impact négatif sur la probabilité de percevoir des pensions de retraite, qui accompagne logiquement l'impact observé sur l'âge moyen de la population adulte. On n'observe pas, en revanche, d'effet robuste sur les revenus moyens. Ces résultats sur la structure des revenus et sur la structure démographique présentent un profil cohérent, et suggèrent une évolution des zones connectées à la fibre vers une structure de population plus jeune et vers une activité non-salariée plus dynamique.

Ceci invite naturellement à examiner les dynamiques immobilières, qui pourraient en théorie refléter une telle évolution démographique. Le taux de vacance des logements diminue de manière causale dans les parcelles raccordées, ce qui est compatible avec l'augmentation du nombre de foyers fiscaux constatée dans les données fiscales ; ainsi que la part des logements proposés à la location. Nous ne constatons pourtant pas d'effet robuste sur la fréquence des transactions immobilières, ou sur les prix immobiliers pratiqués, ni sur la probabilité d'implantation d'une entreprise.

ANNEXE :

VARIANTES DE DATE DE TRANSITION

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

Bien que la méthode des études d'évènements soit très fréquemment employée, les conditions sous lesquelles elle permet d'aboutir à une estimation non-biaisée des effets qu'elle vise à mesurer ne sont pas toujours explicitées. Cette sous-section a pour objet de clarifier les hypothèses d'identification sous-tendant les études d'évènement. Pour étudier sous quelles hypothèses l'estimation de l'équation (3.1) estime l'effet de traitement moyen sans biais, il est utile de suivre la version “document de travail” de Abraham et Sun (2019) et d'introduire une notation en résultat potentiel (*potential outcome*).

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

puis de uns, c'est-à-dire $D_{i,t} < D_{i,s}$ pour $s < t$.

Nous notons $Y_{i,t}^c$ pour indiquer le résultat potentiel dans la période t sous une trajectoire de traitement hypothétique (date de l'événement) c . Soit $Y_{i,t}^B$ pour indiquer le résultat si l'unité i ne reçoit aucun traitement. Il s'agit de la trajectoire de base, c'est-à-dire celle qu'aurait suivi l'unité i en l'absence de traitement.

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

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

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

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

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

Abraham et Sun (2019) montrent que sous trois ensemble d'hypothèses, l'estimation de l'équation (3.1) par moindres carrés ordinaires aboutit à une estimation causalement interprétable de l'effet moyen.

1. La première hypothèse¹⁰ concerne l'homogénéité des effets moyens entre "cohortes" où une cohorte fait référence à l'ensemble des unités qui ont reçu le traitement à une date donnée (par exemple la cohorte c est définie comme : $\{i \in N : t_{i0} = c\}$). Le traitement moyen pour la cohorte c pour une distance au traitement donnée l est défini comme $CATT_l^c = E(\alpha_{i,c+l}^c | t_{i0} = c)$. L'hypothèse

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

s'écrit donc :

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

2. La deuxième hypothèse¹¹ est une généralisation de l'hypothèse de tendance commune entre cohortes. Formellement :

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

où supp est le support de la variable aléatoire.

3. La troisième hypothèse¹² concerne l'absence d'anticipation. Elle revient à supposer que le traitement futur n'a pas d'effet dans le passé. Formellement :

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

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

Dans le cadre de notre application au niveau ville, l'hypothèse d'homogénéité de l'effet du traitement entre cohortes peut être problématique. Les villes traitées en 2015 sont différentes de celles traitées en 2018. Même si l'hypothèse de tendance commune tient entre ces groupes, il est possible que les effets du THD soient plus forts dans les villes traitées en premier. L'hétérogénéité liée à des évolutions structurelles de l'économie au cours de la période est vraisemblablement limitée du fait de la courte durée de la période couverte par l'échantillon.

Une hypothèse plus problématique est l'absence d'effet d'anticipation. En effet,

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

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

recevoir la fibre peut être affecté par l'activisme des élus locaux (par exemple) qui sera plus élevé et efficace dans les zones économiquement dynamiques. Dès lors il n'est pas clair que la présence de tendance pré-traitement (*pre-trends*) reflète l'anticipation de la politique ou bien une violation de l'hypothèse 2 de tendance commune.

RÉFÉRENCES

- Abraham, S. et Sun, L. (2019). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Available at SSRN 3158747*.
- Ahlfeldt, G., Koutroumpis, P., et Valletti, T. (2017). Speed 2.0 : Evaluating access to universal digital highways. *Journal of the European Economic Association*, 15(3), p. 586–625.
- Baker, A., Larcker, D. F., et Wang, C. C. (2021). How much should we trust staggered difference-in-differences estimates? *Available at SSRN 3794018*.
- Bertschek, I., Briglauer, W., Hüschelrath, K., Kauf, B., et Niebel, T. (2015). The economic impacts of broadband internet : A survey. *Review of Network Economics*, 14(4), p. 201–227.
- Borusyak, K. et Jaravel, X. (2017). Revisiting event study designs with an application to the estimation of the marginal propensity to consume. *mimeo*, p. 33.
- Borusyak, K., Jaravel, X., et Spiess, J. (2021). Revisiting event study designs : Robust and efficient estimation. *arXiv preprint arXiv :2108.12419*.
- Callaway, B. et Sant’Anna, P. H. (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.
- Cengiz, D., Dube, A., Lindner, A., et Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3), p. 1405–1454.

- Cerema (2020). *Wébinaire fichiers fonciers : Réponses aux questions du dialogue en ligne. document en ligne.*
- Chay, K. Y. et Greenstone, M. (2005). Does air quality matter? evidence from the housing market. *Journal of political Economy*, 113(2), p. 376–424.
- Czernich, N., Falck, O., Kretschmer, T., et Woessmann, L. (2011). Broadband infrastructure and economic growth. *The Economic Journal*, 121(552), p. 505–532.
- Davis, S. J. et Haltiwanger, J. (1992). Gross job creation, gross job destruction, and employment reallocation. *The Quarterly Journal of Economics*, 107(3), p. 819–863.
- Deshpande, M. et Li, Y. (2019). Who is screened out? application costs and the targeting of disability programs. *American Economic Journal : Economic Policy*, 11(4), p. 213–48.
- Dingel, J. I. et Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, 189, p. 104235.
- Glaeser, E. et Gyourko, J. (2018). The economic implications of housing supply. *Journal of Economic Perspectives*, 32(1), p. 3–30.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), p. 254–277.
- Gross, T., Notowidigdo, M. J., et Wang, J. (2019). The marginal propensity to consume over the business cycle. *American Economic Journal : Macroeconomics*.
- Kim, Y. et Orazem, P. F. (2017). Broadband internet and new firm location decisions in rural areas. *American Journal of Agricultural Economics*, 99(1), p. 1–18.
- Mack, E. A., Anselin, L., et Grubestic, T. H. (2011). The importance of broadband

- provision to knowledge intensive firm location. *Regional Science Policy & Practice*, 3(1), p. 17–35.
- Mack, E. A. et Rey, S. J. (2014). An econometric approach for evaluating the linkages between broadband and knowledge intensive firms. *Telecommunications Policy*, 38(1), p. 105–118.
- Poulhès, M. (2015). Fenêtre sur cour ou chambre avec vue? prix hédoniques de l’immobilier parisien. Technical report.
- Rosen, S. (1974). Hedonic prices and implicit markets : product differentiation in pure competition. *Journal of political economy*, 82(1), p. 34–55.
- Rosenthal, L. (2006). Efficiency and seasonality in the uk housing market, 1991–2001. *Oxford Bulletin of Economics and Statistics*, 68(3), p. 289–317.
- Sun, L. et Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), p. 175–199.
- Vannutelli, S. (2020). From lapdogs to watchdogs : Random auditor assignment and municipal fiscal performance in italy. *Job Market Paper*.



L'Institut des politiques publiques (IPP) est développé dans le cadre d'un partenariat scientifique entre PSE-Ecole d'économie de Paris (PSE) et le Centre de Recherche en Économie et Statistique (CREST). L'IPP vise à promouvoir l'analyse et l'évaluation quantitatives des politiques publiques en s'appuyant sur les méthodes les plus récentes de la recherche en économie.

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



Le Groupe des écoles nationales d'économie et statistique (GENES) est un établissement public d'enseignement supérieur et de recherche. Au sein du GENES, le CREST est un centre de recherche interdisciplinaire spécialisé en méthodes quantitatives appliquées aux sciences sociales regroupant des chercheurs l'ENSAE Paris, de l'ENSAI, du département d'Économie de l'École polytechnique et du CNRS. <http://www.groupe-genes.fr/> – <http://crest.science>

