

POLITECNICO DI TORINO

Laurea Magistrale in Ingegneria Gestionale

Tesi di Laurea Magistrale

ANALISI DELLA DIFFUSIONE DELLA BANDA
ULTRAVELOCE IN ITALIA: STUDIO DELL'IMPATTO
SULL'OCCUPAZIONE



**Politecnico
di Torino**

Relatore

Prof. Carlo Cambini

Candidato

Davide Fontana

A.A. 2022-2023

*Alla mia famiglia, per il supporto fornitomi nella
stesura di questo elaborato*

Sommario

INTRODUZIONE	1
CAPITOLO 1: BANDA LARGA E BANDA ULTRALARGA	3
1.1 DEFINIZIONI E DISTINZIONI	3
1.2 TECNOLOGIE ED INFRASTRUTTURE DI RETE	4
1.2.1 INFRASTRUTTURA A MONTE	5
1.2.2 MODALITA DI CONNESSIONE DELL'ULTIMO MIGLIO	6
1.3 IL QUADRO EUROPEO E ITALIANO	9
1.3.1 I PIANI EUROPEI	9
1.3.2 LA STRATEGIA ITALIANA	9
1.3.3 IL RUOLO DEGLI INVESTIMENTI PRIVATI	12
1.4 COSTI E BENEFICI DELLA BANDA ULTRALARGA	13
1.5 CENNI SULLA LETTERATURA ECONOMICA	14
CAPITOLO 2: DESCRIZIONE DEI DATI	16
2.1 ANALISI DELLA DIFFUSIONE DELLA BANDA ULTRALARGA	16
2.2 ANALISI DELL'OCCUPAZIONE IN ITALIA	23
CAPITOLO 3: ANALISI ECONOMETRICA	27
3.1 IL METODO OLS (ORDINARY LEAST SQUARES)	27
3.2 SPECIFICHE DEI MODELLI UTILIZZATI	31
3.3 IL METODO TWO STAGE LEAST SQUARES	34
3.2.1 DESCRIZIONE DEL METODO	35
3.2.2 SPECIFICAZIONE DEL MODELLO	37
3.2.3 ANALISI DELLO STRUMENTO UTILIZZATO	38
3.2 ANALISI DEI RISULTATI	42
CONCLUSIONI	46
Bibliografia	48

INTRODUZIONE

Le connessioni Internet ad alta velocità stanno assumendo una sempre maggiore rilevanza nella società di oggi.

Durante la pandemia da COVID19, grazie alle connessioni ultraveloci i ricercatori e gli studiosi hanno potuto condividere quasi istantaneamente le loro scoperte scientifiche trovando una risposta al virus in tempi più brevi. Grazie a queste connessioni è stato possibile per molti lavoratori usufruire dello *smart working* evitando lo stop delle attività produttive e un conseguente collasso economico. Un altro vantaggio dell'Internet veloce è stato apportato alle piattaforme audio e video che hanno garantito la continuità didattica per docenti e studenti superando l'ostacolo del distanziamento fisico dovuto alle misure di contenimento della mobilità.

A livello economico, diversi studi hanno constatato che la banda ultralarga è in grado di apportare aumenti di produttività e di crescita economica. Grazie alla digitalizzazione vi sono una maggiore quantità e qualità di dati che contribuiscono ad un maggior controllo e flessibilità del processo produttivo. Tutto ciò crea economie di scala e dunque una riduzione dei costi operativi. Per le imprese italiane, una connessione Internet veloce e stabile rappresenta un asset fondamentale che le permette di rimanere competitive nell'economia globale. Anche per quanto riguarda il mercato del lavoro è stato constatato che la banda ultralarga può introdurre benefici come un aumento nel numero di occupati e nella retribuzione media conseguenza quest'ultima di un aumento della produttività. Inoltre, una connessione ad Internet veloce facilita l'incontro tra domanda e offerta di lavoro, offrendo la possibilità a molti utenti di cercare più facilmente posizioni aperte in qualsiasi città.

In questo lavoro, ci si pone l'obiettivo di verificare, se esiste, una correlazione tra la diffusione della banda ultralarga e il mercato del lavoro. Più in particolare, si vuole indagare se l'aumento dell'occupazione in determinati luoghi sia da attribuire all'introduzione della connessione Internet ad alta velocità e, se così, quantificarne l'impatto.

Il presente elaborato è così suddiviso: nel primo capitolo viene riportata una descrizione generale dell'infrastruttura di rete italiana, degli obiettivi europei per la diffusione della banda ultralarga e una panoramica generale sulla relativa letteratura; nel secondo capitolo viene effettuata un'analisi sul dataset utilizzato relativo alla copertura della banda ultralarga nei comuni italiani e alla rispettiva evoluzione dell'occupazione nel periodo considerato. Infine, nel terzo capitolo viene effettuata un'analisi econometrica per verificare e quantificare l'esistenza di una relazione significativa tra la diffusione della banda ultralarga e il numero di lavoratori.

CAPITOLO 1: BANDA LARGA E BANDA

ULTRALARGA

1.1 DEFINIZIONI E DISTINZIONI

La banda larga o *broadband* nel campo delle telecomunicazioni indica la velocità di trasmissione di dati informativi attraverso la rete Internet ad una velocità superiore a 2Mbit/s.

La banda ultralarga o *ultrabroadband* fa riferimento ad una velocità di connessione effettiva in download di almeno 30Mbit/s. Una rete in grado di fornire tale velocità è nota come NGA o *Next Generation Access*. Reti che possono supportare velocità in download di oltre 100Mbit/s sono definite *NGA-VHCN (Very High Capacity Networks)*.

La definizione di banda larga però non è univoca e non può basarsi unicamente sul concetto di velocità in quanto sarebbe soggetta a rapida obsolescenza dovuta al rapido sviluppo della tecnologia e delle infrastrutture di rete.

L'*ITU* o *International Telecommunications Union* ha stabilito come collegamento a banda larga quello con capacità trasmissiva superiore alla *ISDN PRA Primary Rate Access* che in Europa è di 2Mbit/s mentre negli Stati Uniti è di 1,5Mbit/s. Questa differenza fa sì che una rete può essere considerata broadband in Europa ma non negli Stati Uniti e viceversa creando delle difformità a livello internazionale. L'*OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico)* invece scelse la soglia di 256 kbit/s, ossia la più bassa velocità cui era commercializzato il servizio di accesso attraverso reti di operatori via cavo o le tradizionali reti in rame degli operatori di telecomunicazioni. Il *Communications Committee della Commissione Europea (COCOM)* ha optato per una soglia ancora più bassa, pari a 144 kbit/s. La

Commissione Europea ha fissato nel DAE (Digital Agenda for Europe) il raggiungimento di una connessione a banda larga per tutti i cittadini europei ad una velocità di 30Mbit/s superando dunque la definizione stessa di connessione in banda larga.

Negli Stati Uniti la *Federal Communications Commission* degli Stati Uniti ha innalzato le soglie per una connessione “a banda larga”, passando dai 4 Mbit/s ai 25 Mbit/s di trasmissione in download, e da 1 Mbit/s a 4 Mbit/s in upload.

1.2 TECNOLOGIE ED INFRASTRUTTURE DI RETE

Un' infrastruttura di rete permette ad individui e/o oggetti di comunicare tra loro attraverso uno scambio di dati. Da un punto di vista architettonico un'infrastruttura di rete può essere ripartita in tre livelli cioè la *rete dorsale*, il *miglio intermedio* e l'*ultimo miglio*. L' ultimo miglio è costituito dagli apparati che collegano la centrale all'abitazione. Il miglio intermedio comprende il collegamento delle centrali all'OPB cioè l'infrastruttura che governa il trasporto a livello nazionale. Infine, la rete dorsale comprende i nodi della rete OPB e le connessioni con questi punti.

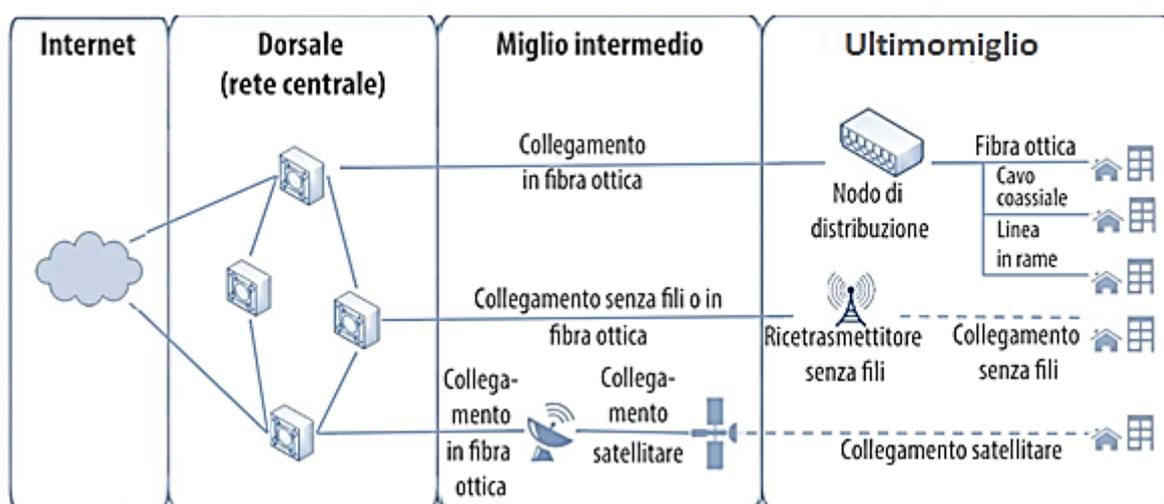


Figura 1.1: Infrastruttura di rete per la connessione a banda larga

1.2.1 INFRASTRUTTURA A MONTE

La trasmissione dei dati avviene attraverso l'*Optical Packet Backbone (OPB)* costituita da 32 nodi e dai collegamenti tra essi. Questi nodi formano un doppio centro stella con un nodo che occupa il posto centrale e mette in collegamento gli altri nodi tra loro. La rete OPB presenta due nodi centrali situati a Milano e a Roma. Per migliorare le proprie prestazioni, la rete OPB utilizza la fibra ottica per collegare i nodi della dorsale e indirizzare il traffico dati verso i *local Central Offices (COs)*. I COs fungono da livello intermedio dell'infrastruttura di telecomunicazione e insieme all'OPB fanno parte della proprietà dell'operatore storico TIM. Nei COs vengono installati specifici dispositivi in grado di convertire i segnali ottici in segnali che possono essere elaborati dall'infrastruttura dell'ultimo miglio.

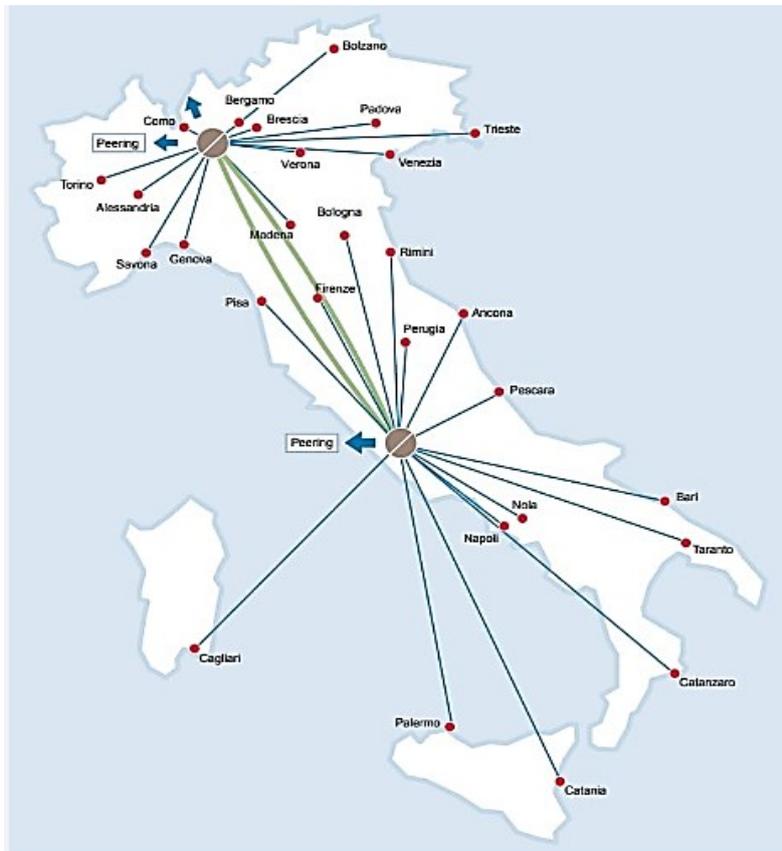


Figura 1.2: Rete dorsale OPB

1.2.2 MODALITA DI CONNESSIONE DELL'ULTIMO

MIGLIO

Con ultimo miglio si intende l'insieme di apparati che consentono il collegamento fra il cliente finale e la più prossima centrale di commutazione. Le tecnologie di connessione possono essere suddivise in:

- Cablate
 - fibra ottica
 - cavo coassiale delle reti televisive
 - rete telefonica tradizionale
- Wireless
 - reti terrestri senza fili (siti di antenne e torri)
 - reti satellitari

In Italia, la connessione a banda larga utilizza in maggior parte i cavi in rame della rete telefonica tradizionale; possiamo definirla una rete di distribuzione “corta”, composta da circa 150'000 armadi di strada con distanze di collegamento limitate e velocità elevate, se confrontate con quelle raggiunte dalla stessa tecnologia in altri Paesi. La disponibilità di tale infrastruttura ha permesso la rapida espansione delle connessioni *DSL (Digital Subscriber Line)* su rame, limitando quella della fibra ottica.

Le tecnologie DSL più affermate sono l'*ADSL (Asymmetric Digital Subscriber Line)* ed il *VDSL2 (Very-high-bit-rate digital subscriber line)*. Queste tecnologie vengono definite asimmetriche dato che il traffico diretto verso il terminale dell'utente (downstream) transita a una velocità maggiore rispetto al traffico in direzione opposta (upstream). Tale asimmetria,

utilizzata per destinare parte dello spettro inferiore ai servizi vocali tradizionali, ben si presta per servizi di navigazione web, che utilizzano principalmente il traffico downstream.

La rete di distribuzione in rame è composta da due tratte:

- Rete Primaria: tratto di rete in rame compreso tra la centrale ed appositi armadi situati in strada, cioè gli armadi ripartilinea;
- Rete Secondaria: tratto di rete in rame compreso tra gli armadi e l'abitazione del cliente;

La Tecnologia ADSL utilizza sia la Rete Primaria che la Rete Secondaria invece la tecnologia VDSL2 solo la Rete Secondaria in rame mentre fa uso di cavi in fibra ottica tra armadi ripartilinea e centrali. In termini di architettura di rete, le soluzioni basate sull'ADSL si chiamano *Fiber To The Exchange (FTTE)*, ossia Fibra sino alla Centrale; le soluzioni che utilizzano il VDSL2 si chiamano *Fiber To The Cabinet (FTTCab)*, ossia Fibra sino all'armadio che divide la Rete Primaria in rame dalla Rete Secondaria.

Queste ultime necessitano di investimenti per la diffusione della fibra ottica sino agli armadi. Le velocità massime dell'ADSL sono circa 25Mbit/s in downstream e 1Mbit/s in upstream mentre quelle della tecnologia VDSL2 sono circa 200Mbit/s in downstream e 50Mbit/s in upstream.

La connessione a banda larga con collegamenti in fibra ottica funziona con una modalità di trasporto diversa rispetto all'ADSL: invece di trasferire segnali elettrici, trasmette impulsi luminosi. Perciò la fibra ottica, non essendo influenzata dal deterioramento del segnale all'aumentare della lunghezza del cavo, garantisce un'elevata qualità della connessione anche nelle zone rurali, contraddistinte da case sparse, mediamente più distanti dalla centrale telefonica e, dunque, finora penalizzate.

Per un operatore di telecomunicazioni, scegliere la fibra ottica significa effettuare una scelta importante, dal momento che viene effettuata una sostituzione, parziale o totale, della rete di distribuzione. Un investimento importante che permette una crescita significativa delle prestazioni, dal momento che la fibra ottica permette velocità più elevate di un ordine di grandezza rispetto alle tecnologie in rame. La velocità in upstream è pari a quella in downstream e può raggiungere un valore massimo di 1 Gbit/s, in base al punto in cui arriva la fibra ottica. Le soluzioni completamente in fibra ottica *FTTH (Fiber to the Home)*, utilizzano la fibra ottica fino all'abitazione del cliente finale e sono caratterizzate da velocità superiori rispetto alle soluzioni che utilizzano il rame per effettuare il collegamento dall'armadio di distribuzione all'abitazione.

Con *rete di accesso di nuova generazione (NGAN)* si intende un'infrastruttura in grado di garantire offerte e servizi che viaggiano a velocità molto superiori rispetto ai tradizionali accessi a banda larga attualmente disponibili. Ad oggi, grazie all'evoluzione delle tecnologie trasmissive sulla rete di accesso tradizionale, si possono sfruttare le grandi potenzialità della rete secondaria in rame e si sta introducendo gradualmente la fibra in rete di accesso, tanto che oggi si definiscono reti NGAN anche le soluzioni che utilizzano la fibra fino all'armadio ripartilinea per poi proseguire in rame fino all'abitazione (FTTCab).

Si può affermare che ogni operatore decide la propria architettura tenendo conto della tipologia di aree da servire, della densità abitativa, della tipologia di clientela, della disponibilità di infrastrutture ottiche o di canalizzazioni adatte all'uso, dell'impatto urbanistico. Questi aspetti sono, come ci si può aspettare, correlati alle strategie d'investimento ed al contesto competitivo e regolatorio.

1.3 IL QUADRO EUROPEO E ITALIANO

1.3.1 I PIANI EUROPEI

La Commissione Europea riconosce l'importanza di garantire a tutti gli Stati membri una migliore copertura di rete considerando periodicamente degli obiettivi da raggiungere.

Con la pubblicazione avvenuta nel marzo 2020 della *Comunicazione COM(2021) 118 final "Bussola digitale 2030: la via europea per il decennio digitale"* la Commissione Europea ha delineato gli obiettivi di connettività per il 2030 tra cui rientrano:

- connettività di almeno 1 Gbps per tutte le famiglie situate nel territorio Europeo
- copertura 5G in tutte le aree popolate

Nel 2016, invece, con la *Comunicazione COM(2016) 587 final "Connettività per un mercato unico digitale competitivo: verso una società dei Gigabit europea"* la Commissione europea ha fissato gli obiettivi per il 2025:

- connettività di almeno 1 Gbps per scuole, biblioteche e pubblici uffici
- connettività di almeno 100 Mbps, espandibile fino a Gigabit, per tutte le famiglie europee
- copertura 5G senza interruzioni in tutte le aree urbane e lungo i principali assi di trasporto terrestre

1.3.2 LA STRATEGIA ITALIANA

Per soddisfare i piani della Commissione Europea per il 2020, il Governo italiano ha varato il 3 Marzo 2015 la *Strategia Italiana per la Banda Ultralarga*, contenente i seguenti punti programmatici:

- Copertura dell'intera popolazione nazionale con connettività superiore ai 30 Mbps
- Copertura fino all'85% della popolazione con una connettività di almeno 100 Mbps al fine di aver sottoscritti contratti di questo tipo per almeno il 50% della popolazione residente in Italia

Per attuare la Strategia sopra definita, dato che il mercato da solo non avrebbe potuto conseguire gli obiettivi prefissati per il 2020, è stato elaborato il *Piano di investimento per la diffusione della banda ultralarga*, contenente gli interventi pubblici necessari a favorire gli investimenti sull'infrastruttura di rete e ad ampliare lo sviluppo della banda ultralarga sia dal lato dell'offerta sia dal lato della domanda.

Riferendosi all'infrastruttura di rete, il Piano ha adottato una nuova modalità di aggregazione territoriale, costituita da 94 645 aree più piccole anziché sui comuni. Questa divisione è dovuta al fatto che la diffusione della copertura di rete non avviene in maniera uniforme sull'intero territorio comunale; diventa quindi necessario suddividere il comune in zone coperte dall'infrastruttura di rete e zone ancora da coprire.

Per quanto riguarda la qualità delle connessioni, si sono individuati quattro cluster, a seconda del livello di intervento pubblico reso necessario per il conseguimento dell'obiettivo.

- *Cluster A - aree redditizie*: questa è l'area più adatta al conseguimento dell'obiettivo di attuare reti ultraveloci a 100 Mbps entro il 2020. Comprende le 15 città più abitate d'Italia e le principali aree industriali. Costituisce il 15% della popolazione nazionale, cioè circa 9,4 milioni di abitanti
- *Cluster B - aree per le quali non è previsto un investimento a 100 Mbps*: si riferisce alle aree in cui gli operatori privati intendono realizzare connessioni a 30 Mbps; in queste

aree, senza l'intervento pubblico le condizioni di mercato non soddisfano i ritorni minimi necessari agli operatori che effettuano investimenti per una connessione a 100 Mbps. Comprende 1120 comuni, ovvero circa 28,2 milioni di abitanti. È suddiviso ulteriormente in due sotto cluster, a seconda che sia stato portato avanti o no un intervento pubblico per lo sviluppo della connettività ad almeno 30 Mbps.

- *Cluster C - aree marginali*: In queste aree, grazie a un sostegno statale, gli operatori potrebbero effettuare investimenti in connessioni superiori ai 100Mbps. Sono circa 2.650 comuni e alcune aree rurali senza reti a più di 30 Mbps. Vi risiedono circa 15,7 milioni di persone.
- *Cluster D - aree a fallimento di mercato*: A causa della loro bassa densità abitativa e della dislocazione frastagliata sul territorio, solo l'intervento pubblico diretto può permettere alla popolazione residente di avere un servizio di connettività a più di 30 Mbps. Il cluster D conta i restanti 4.300 comuni, cioè circa il 15% della popolazione.

Un'ulteriore classificazione del territorio italiano consta nella sua suddivisione in tre tipologie di aree in base alle scelte di investimento delle società di telecomunicazione:

- *"Aree nere"*: sono le aree nelle quali vi sono almeno due operatori di mercato che hanno pianificato investimenti per la realizzazione della banda ultralarga nel successivo triennio (tali aree corrispondono grosso modo a quelle del cluster A)
- *"Aree bianche"*: sono quelle nelle quali non c'è nessun operatore di mercato che, nei propri piani di investimento dei successivi tre anni, sia intenzionato ad investire per la realizzazione dell'infrastruttura (corrispondenti in sostanza ai cluster C e D)
- *"Aree grigie"*: sono aree nelle quali vi è un solo operatore commerciale interessato ad investire per la realizzazione dell'infrastruttura (corrispondenti al cluster B)

Gli investimenti economici statali, atti a favorire il raggiungimento delle aree a fallimento di mercato con reti in fibra ottica, sono stati affidati tramite tre bandi di gara aggiudicatisi sempre dalla società *Open Fiber S.p.a.*

Nel 2021 sono state avviate le consultazioni in merito al piano di intervento “*Italia a 1 Giga*” approvato il 27 luglio dello stesso anno dal *Comitato Interministeriale per la transizione digitale (CITD)*. Con questo piano il Governo italiano fornirà una connettività di almeno 1 Gbit/s in download e di 200 Mbit/s in upload alle unità immobiliari che, a seguito della mappatura delle infrastrutture presenti o pianificate al 2026 dagli operatori di mercato, sono risultate non coperte da almeno una rete in grado di fornire in maniera affidabile velocità di connessione in download pari o superiori a 300 Mbit/s. Ogni stazione appaltante configurerà la procedura di gara individuando un numero di *lotti* che, tenuto conto della specifica struttura del mercato, risulti idoneo a garantire la corretta dinamica concorrenziale in fase di gara e l’efficienza dell’intervento pubblico. I lotti sono costituiti generalmente da regioni o da un insieme di regioni confinanti tra loro. Al 24 maggio 2022 risultavano assegnati 14 dei 15 lotti disponibili. Il 28 giugno è stato assegnato l’ultimo lotto comprendente le Province autonome di Trento e Bolzano. L’importo del finanziamento per attuare questo piano è stato di circa 3,4 miliardi di Euro.

1.3.3 IL RUOLO DEGLI INVESTIMENTI PRIVATI

La diffusione della banda ultralarga, oltre che dagli investimenti pubblici, è stata possibile grazie agli investimenti dei due principali operatori nel campo delle telecomunicazioni, cioè *Tim* e *Open Fiber*.

Le prime infrastrutture di rete a banda ultralarga sono state realizzate dall'operatore incumbent Tim che ha potuto trarre vantaggio dalla sua preesistente infrastruttura, utilizzata per servizi vocali e ADSL, raggiungendo nel 2015 quasi tutti i comuni italiani con una tecnologia mista fibra-rame.

Open Fiber, invece, è un operatore *wholesale* attivo dal 2017 di proprietà di *Cassa Depositi e Prestiti* che utilizza un'infrastruttura basata esclusivamente sulla tecnologia FTTH.

1.4 COSTI E BENEFICI DELLA BANDA ULTRALARGA

La banda ultralarga è in grado di generare effetti positivi diretti e indiretti sulla attività economiche. I benefici diretti si traducono in un aumento dell'occupazione e delle attività economiche per la realizzazione di un'infrastruttura di rete. I benefici indiretti consistono invece nell'aumento della produttività conseguente all'adozione di una nuova tecnologia.

La possibilità garantita dalle connessioni veloci di inviare e ricevere dati e documenti fa sì che il costo di input per le imprese diminuisca. Considerando la teoria dell'impresa tale diminuzione provocherebbe una riduzione o un aumento della domanda di lavoro a seconda che il lavoro costituisca un sostituto o un complemento per queste tecnologie. Ciò potrebbe far sì che l'attività di alcune figure del personale aziendale venga sostituita dalle nuove tecnologie. Dall'altro lato, però, la banda ultralarga fa sì che la produttività delle imprese aumenti generando un conseguente aumento dell'occupazione e dei salari medi.

Ciò dipende dai diversi settori considerati. Vi saranno aziende che considereranno il lavoro complementare alla tecnologia a banda larga ricercando quindi figure maggiormente specializzate accentuando quindi il divario tra personale qualificato e non qualificato.

Grazie alle connessioni veloci, i clienti possono rivolgersi con maggiore facilità alle imprese online. Questo crea maggiore concorrenza per le imprese locali aumentando il loro rischio di fallimento. D'altro canto, le imprese localizzate in luoghi più piccoli o isolati possono più facilmente venire in contatto con mercati più grandi ampliando le loro possibilità di guadagno.

1.5 CENNI SULLA LETTERATURA ECONOMICA

Diversi studi hanno trovato una correlazione significativa tra la copertura della banda larga e la crescita dell'occupazione mentre altri non hanno osservato alcun miglioramento nel numero di occupati dalla disponibilità di questa tecnologia o da un aumento della velocità di connessione. La prima tesi è sostenuta da *Kolko (2012)* basandosi su osservazioni relative al periodo 1999-2006 constatando che grazie alla banda larga si verifica un aumento dell'occupazione anche quando la crescita della popolazione viene controllata. Tuttavia, Kolko ha rilevato che le aree con una più rapida espansione della banda larga non hanno visto aumenti del tasso di occupazione o della retribuzione media. Ciò può essere spiegato da un aumento del trasferimento o del pendolarismo dei lavoratori, incentivato dalla maggiore occupazione, verso aree dove l'offerta di lavoro è maggiore. *Atasoy (2013)* conferma l'effetto positivo, significativo e causale della banda larga sul tasso di occupazione rilevando che l'incremento non è provocato da un maggior numero di imprese ma da un aumento della forza lavoro in quelle già esistenti. Secondo Atasoy però questo impatto positivo coinvolge principalmente i lavoratori laureati maggiormente richiesti nelle contee americane con industrie alla ricerca di lavoratori altamente qualificati e istruiti.

L'analisi effettuata da *Bai (2016)* invece, considerando i dati relativi al periodo 2011-2014, non ha evidenziato un aumento dell'occupazione a seguito dell'introduzione della banda larga. Alla stessa conclusione è giunto *Ford (2018)* sottolineando che, a seguito di un aumento della velocità di connessione da 10 a 25 Mbps, non si rileva un aumento significativo dell'occupazione. Lo studio condotto da *Czernich (2014)* si è servito di dati relativi ai comuni tedeschi prendendo in considerazione la disponibilità o meno della DSL per valutare l'effetto della banda larga sui tassi di disoccupazione. Utilizzando dapprima un metodo dinamico di regressione ai minimi quadrati, l'autore ha constatato che la disponibilità della banda larga influisce negativamente sul tasso di disoccupazione; utilizzando invece un approccio a variabili strumentali per tenere in considerazione problematiche di variabili omesse e causalità inversa, Czernich ha potuto constatare che una maggiore disponibilità di banda larga non produce causalmente un minore tasso di disoccupazione.

E da notare che gli effetti della banda larga differiscono a seconda del luogo e dei settori economici. *Kolko (2012)* ha notato che la correlazione positiva tra disponibilità della banda larga e crescita dell'occupazione è più rilevante nelle industrie che operano nelle tecnologie dell'informazione. Gli studi effettuati da *Ivus e Boland (2015)* hanno confermato questa ipotesi considerando dati di diffusione della banda larga in Canada nel periodo 1997-2011. In particolare, hanno evidenziato un'influenza positiva della banda larga sulla crescita dell'occupazione principalmente nel settore dei servizi mentre nessun effetto è stato rilevato nel settore manifatturiero. Inoltre, hanno rilevato che l'effetto positivo si manifesta solamente nelle aree rurali. Analogamente, *Atasoy (2013)* conferma che le aree rurali traggono maggiori benefici delle aree urbane dall'aumento della disponibilità della banda larga anche se questo vantaggio è piuttosto modesto.

CAPITOLO 2: DESCRIZIONE DEI DATI

In questo capitolo, verrà svolta un'analisi dei dati relativi alla diffusione della banda ultralarga in Italia; dopo di che si procederà ad analizzare i dati sull'occupazione relativi al periodo 2019-2021.

2.1 ANALISI DELLA DIFFUSIONE DELLA BANDA ULTRALARGA

L'analisi in questione è stata effettuata su un dataset comprendente 7809 comuni italiani in cui per ogni anno, dal 2019 al 2021, è verificata la disponibilità di una connessione ultraveloce grazie ai dati forniti da *Infratel*. Nonostante i comuni italiani siano 7904, questo dataset può considerarsi rappresentativo dell'intero territorio nazionale dato che le osservazioni sono in numero elevato. Sono stati esclusi alcuni comuni per la mancanza di dati provocata dalle variazioni territoriali e amministrative avvenute come il cambio delle circoscrizioni territoriali, l'istituzione di nuove, i cambi di denominazione e le soppressioni.

Dai dati in possesso, si nota che i comuni coperti dalla banda ultralarga sono passati da circa il 53% del campione considerato nel 2019 al 72% nel 2021 come mostrato nella figura 2.1. L'incremento maggiore si è verificato dal 2020 al 2021 dove la crescita è stata di circa il 12%. Nel periodo precedente 2019-2020, l'aumento è stato di circa il 7%. Questo maggiore incremento è in parte dovuto al fatto che, a seguito della pandemia, parecchi lavoratori hanno proseguito le proprie attività lavorative da remoto e dunque sempre più amministratori locali hanno provveduto a portare ai loro cittadini una connessione veloce per assecondare le loro esigenze.

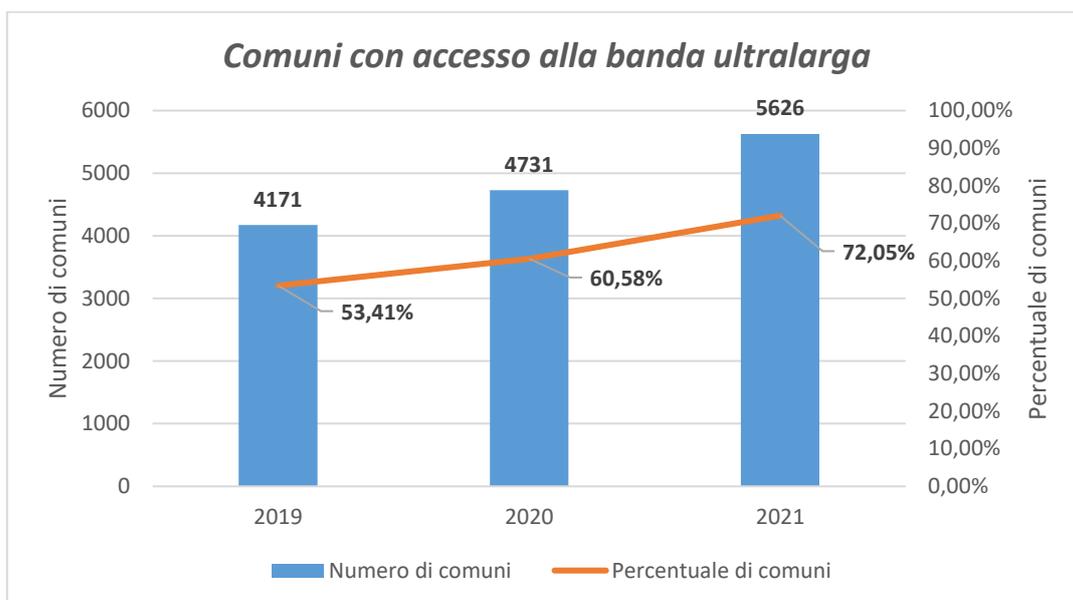


Figura 2.1: Diffusione della banda ultralarga nei comuni italiani

Per quanto riguarda la popolazione coperta, la figura 2.2 mostra che ben il 93% dei cittadini può servirsi di una connessione ultraveloce nel 2021. Ciò dimostra come gli operatori di telecomunicazione si siano concentrati maggiormente dapprima sulle aree a maggiore densità abitativa soprattutto per motivi di profittabilità. Rimangono da coprire dunque comuni sempre meno popolosi.

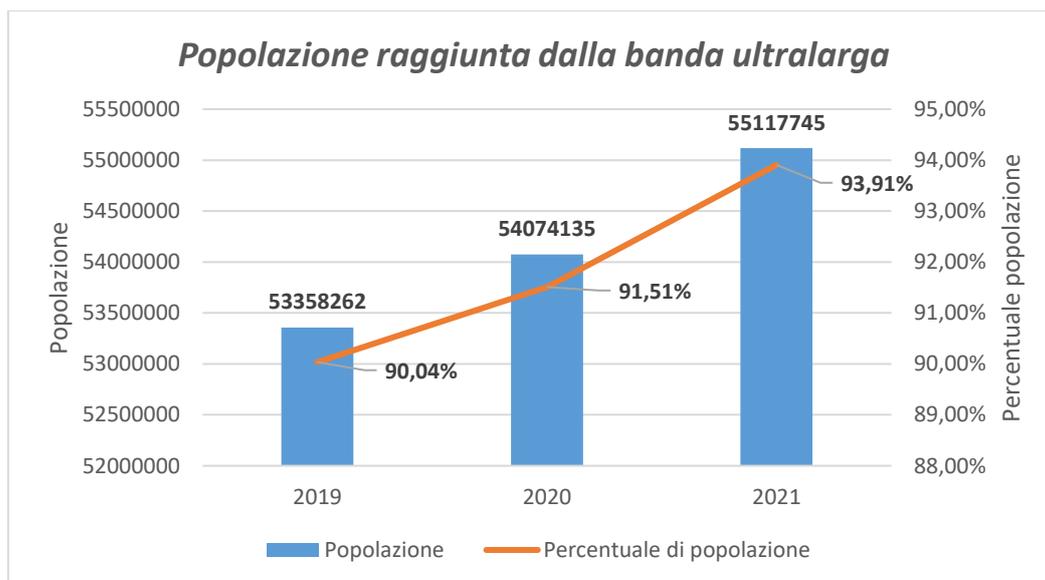


Figura 2.2: Diffusione della banda ultralarga sulla popolazione italiana

E da notare come, rispetto al resto d'Europa, l'Italia sia in ritardo in termini di connettività. Questo è dovuto al fatto che nelle grandi città, cioè quelle con più di 250000 abitanti, risiede solo il 16% della popolazione italiana mentre il 50% risiede in centri con meno di 25000 abitanti. Negli altri stati, la situazione è diversa con molti più abitanti concentrati nei grandi centri urbani. Roma la città più popolosa d'Italia, secondo i dati *Eurostat* del 2014, conta appena il 4% della popolazione nazionale. Siccome nelle grandi città, i costi per la realizzazione di reti NGA sono inferiori per singola unità immobiliare, l'Italia si trova in una situazione di svantaggio rispetto ad altri Paesi.

Analisi a livello regionale

Per quanto riguarda la copertura della banda ultralarga nelle diverse regioni italiane si nota dalla figura 2.3 che la regione con il maggior numero di comuni coperti nel 2021 è la Lombardia con 1062 comuni seguita da Piemonte e Campania con 595 e 477 comuni rispettivamente.

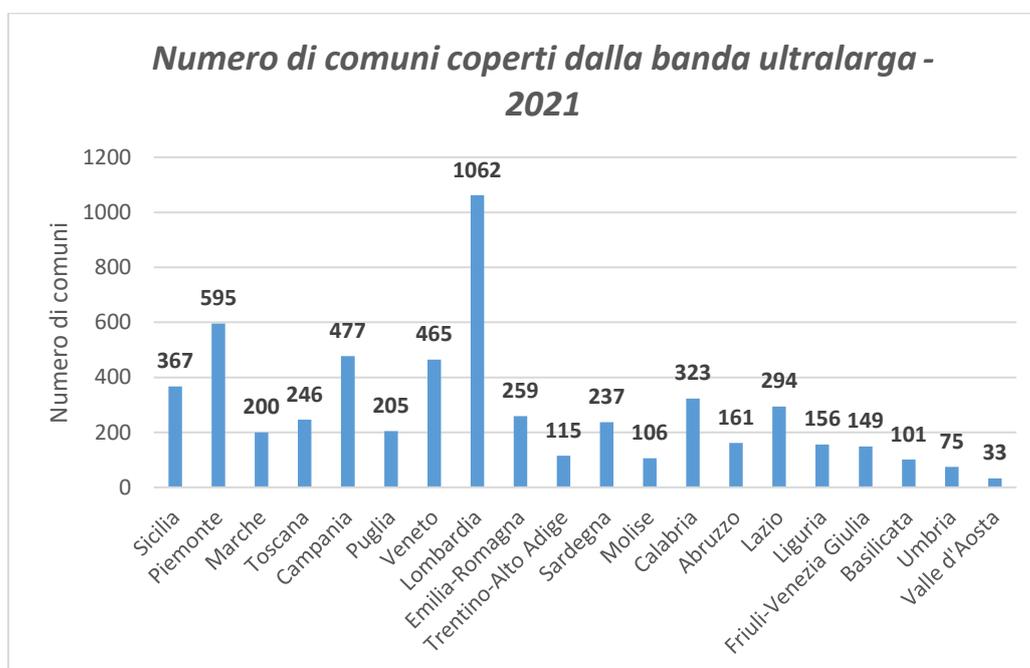


Figura 2.3: Diffusione della banda ultralarga per regione a fine 2021

Rapportando il numero di comuni coperti al totale per ogni regione è possibile definire il livello regionale di diffusione delle NGA. La figura 2.4 mostra che a fine 2021 ben 12 regioni su 20 hanno una copertura che va dal 75% al 100%, 6 regioni una copertura dal 50% al 75% e 2 regioni una copertura che va dal 25% al 50%. La Lombardia e il Piemonte hanno una copertura rispettivamente del 71,18% e del 50,81% però contano un numero di comuni decisamente più elevato rispetto alle altre regioni; in particolare il Piemonte ne ha 1171 e la Lombardia 1492 mentre la Toscana e la Sicilia ai primi due posti ne contano 259 e 390 rispettivamente.

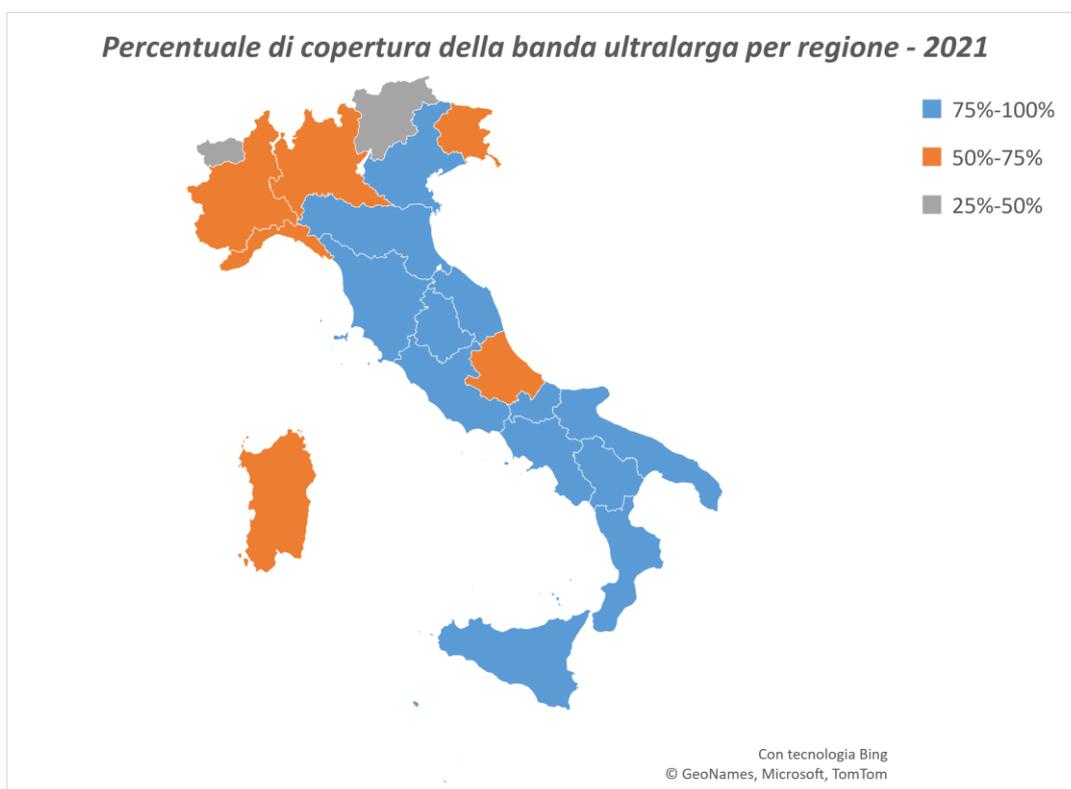


Figura 2.4: Percentuale di copertura della banda ultralarga nelle regioni italiane a fine 2021

I dati sono stati poi aggregati per area geografica per capire in quale parte d'Italia si sono concentrati maggiormente gli investimenti in banda ultralarga. La figura 2.5 mostra l'evoluzione della copertura della banda ultralarga nelle aree considerate nel periodo 2019-2021.

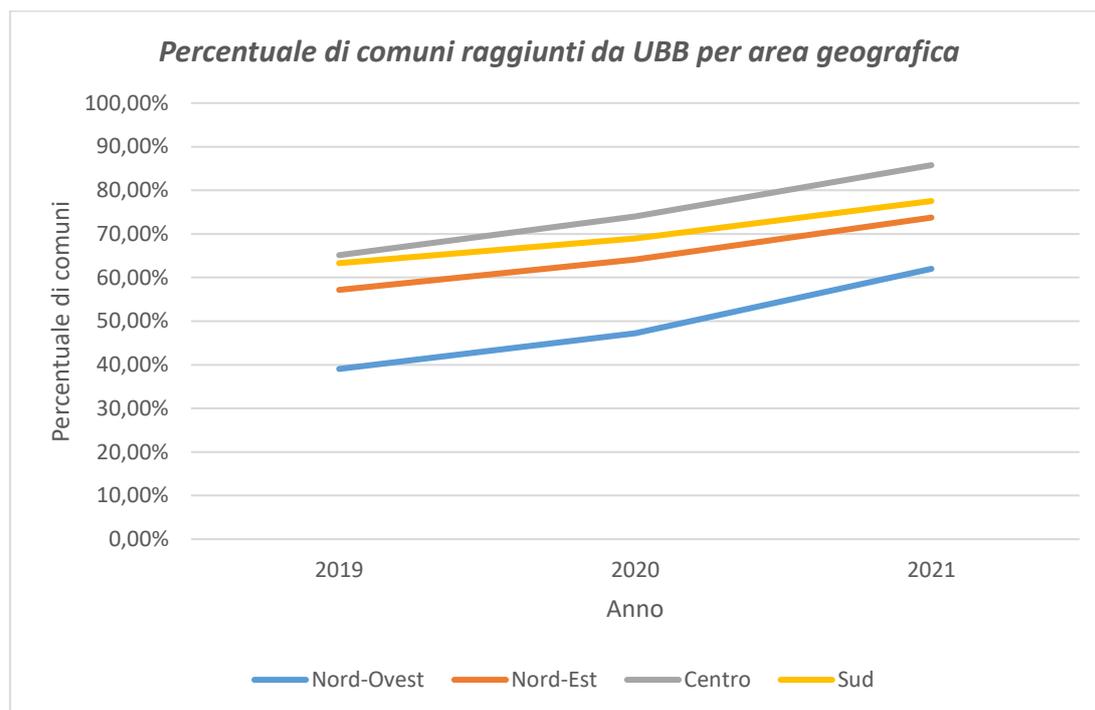


Figura 2.5: Percentuale di comuni coperti rispetto al totale della banda ultralarga per area geografica

Dalla figura si evince che la copertura maggiore della banda ultralarga è nel Centro Italia. Nell'intervallo 2019-2020 la maggiore crescita nella copertura si è registrata nel Centro Italia pari all'8,9% mentre nel periodo 2020-2021 si è registrata al Nord-Ovest con un valore del 14,81%.

Analisi della copertura con tecnologia FTTH

Come si può notare dalla figura 2.6 la diffusione della tecnologia FTTH risulta ancora piuttosto contenuta a fine 2021. Nonostante il trend in crescita, solamente 802 comuni in Italia possono usufruire di questa tecnologia pari a circa il 10% del campione analizzato.

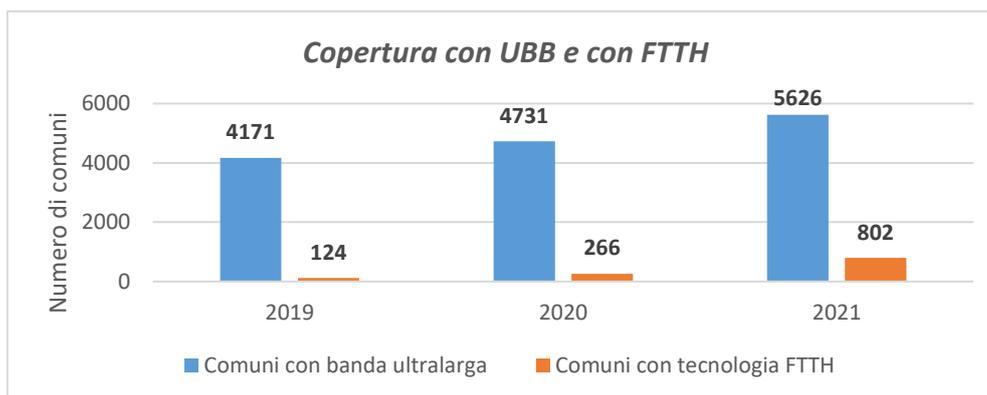


Figura 2.6: Numero di comuni coperti con UBB e con tecnologia FTTH durante il triennio 2019-2021

Analizzando le diverse regioni italiane, il maggior numero di comuni coperti da FTTH a fine 2021 si ha in Campania con 157 comuni seguita da Lombardia con 153 e Piemonte con 101 comuni. Si evince dunque che gli operatori preferiscono adoperare i meno costosi protocolli FTTC con fibra mista a rame in cui la fibra arriva fino agli armadi di strada mentre la distanza tra armadio e abitazione è realizzata in rame.

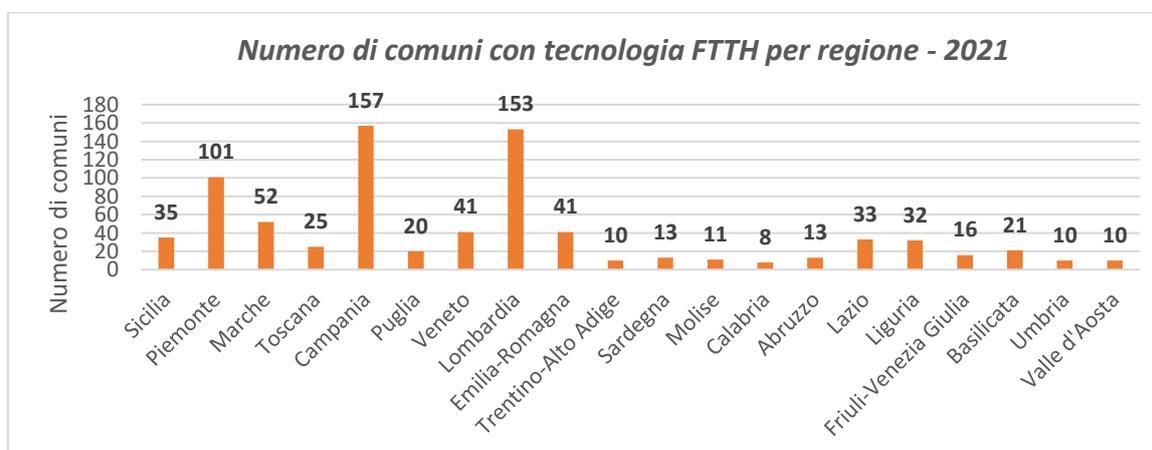


Figura 2.7: Diffusione della tecnologia FTTH per regione italiana

Analisi della copertura per operatore

Per quanto riguarda gli operatori, dalla figura 2.8 si può notare che nei 3 anni considerati la maggior parte dei comuni sia coperta con la rete di *TIM*, cioè dell'operatore incumbent che a differenza di *Open Fiber* fa uso della tecnologia FTTC. E da notare però la rilevante crescita di comuni coperti solamente da *Open Fiber* che nel periodo 2020-2021 sono passati da 90 a 399 con un aumento del 343% rispetto all'anno precedente.

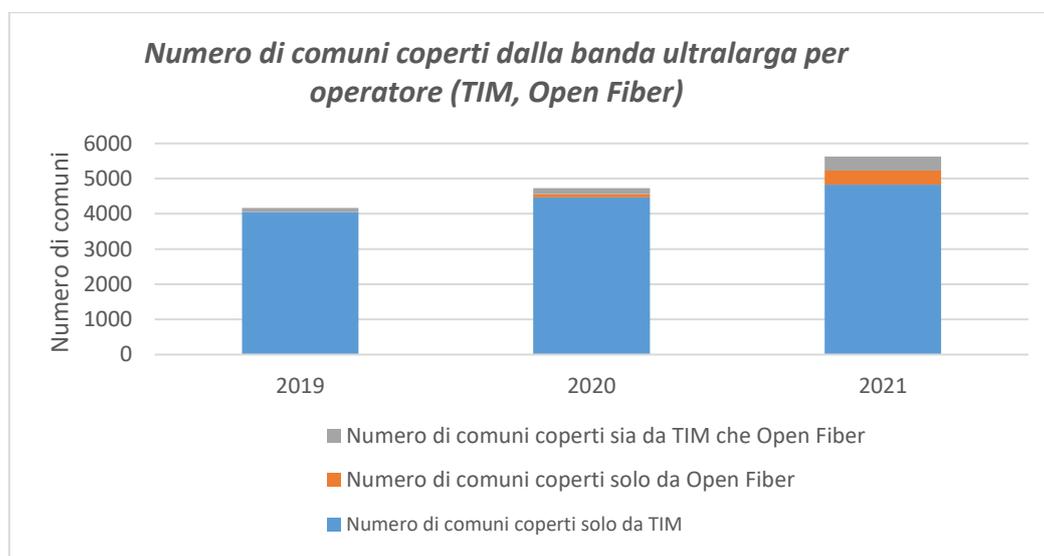


Figura 2.8: Diffusione della banda ultralarga per operatore

2.2 ANALISI DELL'OCCUPAZIONE IN ITALIA

Il dataset utilizzato per effettuare l'analisi dell'occupazione italiana è quello messo a disposizione dal *Ministero dell'Economia e delle Finanze (MEF)* per il periodo 2019-2021. È stato monitorato l'andamento dell'occupazione degli stessi comuni utilizzati per rilevare la copertura della banda ultralarga. Come già detto in precedenza, si considera che questi comuni, rappresentando la quasi totalità di quelli italiani, siano rappresentativi dell'intero territorio nazionale.

Come possibile vedere dalla figura 2.9, nel triennio analizzato l'occupazione subisce un calo nel 2020 a causa dei risaputi effetti provocati dalla pandemia da COVID 19. Tuttavia, l'adozione di misure di contenimento del virus e il ricorso allo smart working da parte di numerose aziende ha fatto sì che il calo degli occupati fosse di solo l'1,08% rispetto all'anno precedente. Nel 2021, grazie alla scoperta del vaccino e all'entrata in vigore del *PNNR* cioè del *Piano Nazionale di Ripresa e Resilienza* c'è stato un rimbalzo nel numero di occupati che ha toccato quota 22900474 con una crescita del 2% rispetto al 2020 spinto anche dall'aumento dei contratti a termine.

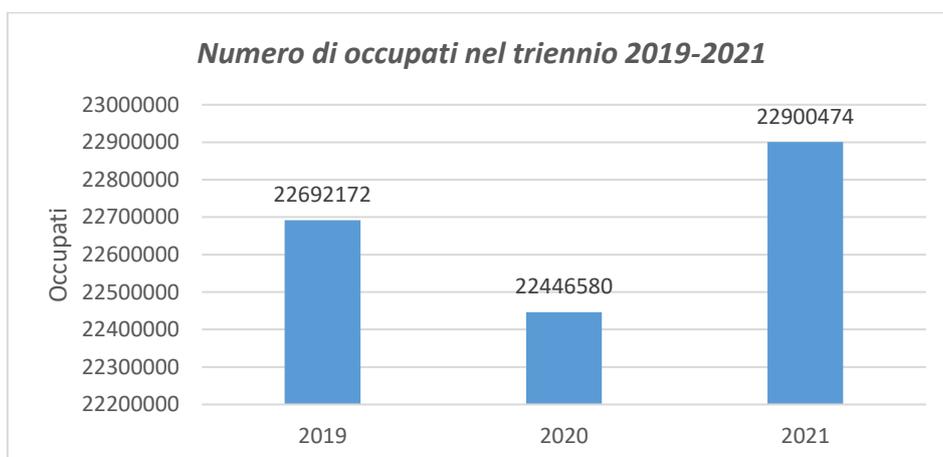


Figura 2.9: Evoluzione nel numero degli occupati nel triennio 2019-2021

Per quanto riguarda il tasso di occupazione, definito come il rapporto tra gli occupati e la popolazione in età lavorativa, si può notare in figura 2.10 una flessione dello 0,49% nel periodo 2019-2020 e un aumento dello 0,91% in quello 2020-2021.

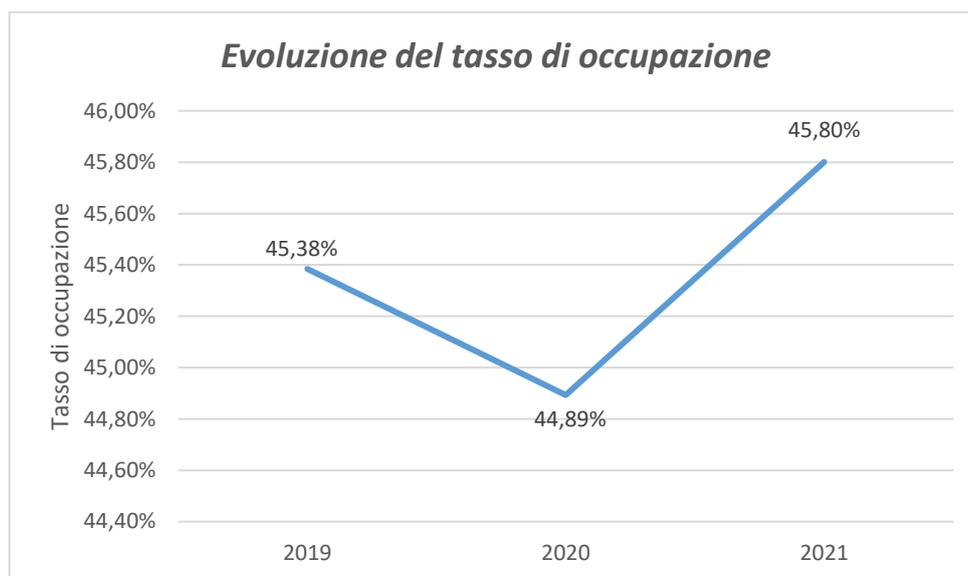


Figura 2.10: Evoluzione del tasso di occupazione nel triennio 2019-2021

Nella tabella 2.1 è invece riportato il saldo assunzioni-licenziamenti per macroarea italiana. Come è possibile notare, il calo dell'occupazione durante la pandemia del 2020 ha colpito in misura maggiore il Nord-Est che ha visto l'occupazione ridursi dell'1,4% rispetto al 2019. La ripresa dell'occupazione nel 2021 ha coinvolto tutte le macroaree italiane ed è stata più spinta al Centro Italia con un aumento del 2,39% rispetto all'anno precedente.

		Anno		
		2019	2020	2021
Area	Nord-Ovest	1,70%	-1,09%	1,80%
	Nord-Est	1,51%	-1,41%	2,12%
	Centro	1,06%	-1,18%	2,39%
	Sud	1,58%	-0,76%	1,91%

Tabella 2.1: Variazione percentuale del saldo assunzioni-licenziamenti suddiviso per Macro Area

Facendo riferimento all'ultimo anno analizzato, cioè il 2021, la figura 2.11 mostra come sono ripartiti gli occupati per area geografica. E da notare come per via della pandemia del 2020 che colpì in misura maggiore il Nord Italia ci sia stato un aumento degli occupati al Sud che a fine 2021 ha registrato una percentuale di occupati sul totale del 29,08% superiore a quello delle altre aree geografiche considerate. Questa percentuale è influenzata dal notevole aumento dell'adozione dello smart working, anche in forma permanente, e da altri fattori quali ad esempio la minore diffusione del contagio rispetto ad altre parti d'Italia e il caro affitti che si sta manifestando specie al Nord. E altresì da considerare il fatto che il Sud Italia ha il numero maggiore di regioni rispetto alle altre macroaree considerate.

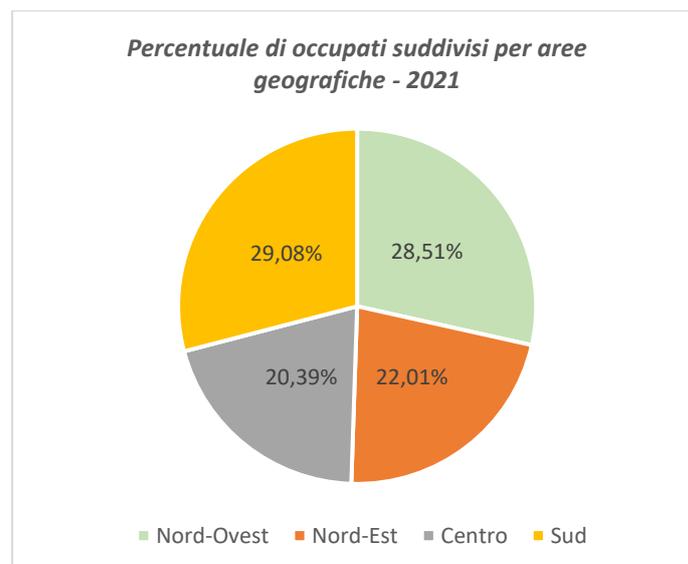


Figura 2.11: Percentuale di occupati sul totale suddivisi per aree geografiche

Passando ad un'analisi a livello regionale, si nota però come la Lombardia rimanga la regione con la percentuale di occupati sul totale maggiore pari al 18,35%. Questo dato è, come si può immaginare, condizionato dalla dislocazione delle imprese sul territorio nazionale. Secondo l'ISTAT, la Lombardia possiede il primato nazionale nel numero di imprese con quasi 800 mila aziende sparse per la regione. E' seguita dal Veneto e dal Lazio che possiedono singolarmente circa 400 mila imprese. Anche per quanto riguarda il numero di occupati Lazio e Veneto si posizionano al secondo e terzo posto dopo la Lombardia a conferma del fatto che esiste una forte correlazione tra il numero di aziende e quello degli occupati. All'ultimo posto per numero di occupati ci sono il Molise e la Valle d'Aosta con una percentuale di occupati dello 0,44% e dello 0,24% rispettivamente. Queste percentuali sono però giustificate dal momento che queste due regioni possiedono un numero di residenti più basso rispetto alle altre regioni italiane.

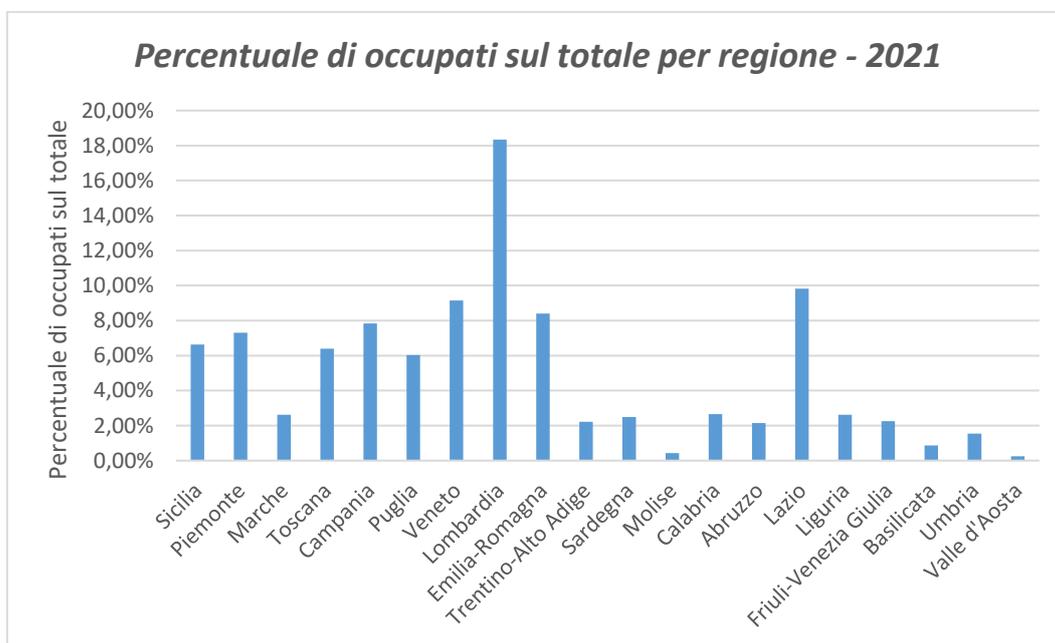


Figura 2.12: Percentuale di occupati sul totale suddivisi per regione

CAPITOLO 3: ANALISI ECONOMETRICA

Questo capitolo ha come obiettivo l'individuazione e la quantificazione dell'impatto della banda ultralarga sul mercato del lavoro, in particolare sull'occupazione, utilizzando modelli econometrici. L'analisi econometrica effettua una stima dei parametri di una funzione che lega due o più variabili con i dati campionari che si hanno a disposizione. In questo studio, si è utilizzato un set di dati panel contenente il numero di occupati per anno dal 2019 al 2021 e i dati scaglionati relativi all'introduzione della banda ultralarga sempre nello stesso intervallo temporale.

3.1 IL METODO OLS (ORDINARY LEAST SQUARES)

Il metodo dei minimi quadrati ordinari o *Ordinary Least Squares* è il metodo più diffuso per valutare l'effetto causale della variazione di una variabile indipendente sulla variabile dipendente. Questo modello viene impiegato nei modelli di regressione lineare e per descriverlo si fa dapprima riferimento al modello più semplice sotto illustrato:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (1)$$

In cui:

- i rappresenta l'osservazione generica (X_i, Y_i) delle n osservazioni che fanno parte del campione analizzato
- X è la variabile indipendente anche nota come regressore
- Y è la variabile dipendente

- β_0 è l'intercetta e β_1 la pendenza della retta stimata. Questi due parametri costituiscono i parametri da stimare. Più in particolare β_1 costituisce l'effetto su Y che si verifica a seguito di una variazione unitaria di X
- ε è il termine d'errore e contiene tutti i fattori diversi da X che vanno ad influenzare Y meglio noti come *variabili omesse*

Tra le infinite coppie (β_0, β_1) lo stimatore ai minimi quadrati seleziona quella che riduce al minimo la somma dei quadrati dei residui. Il residuo, cioè ε , è la differenza tra il valore reale di y e quello predetto \hat{y} dalla retta di regressione calcolata ($\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$). La formula relativa allo stimatore OLS è la seguente:

$$\min \hat{\beta}_0 \hat{\beta}_1 \sum_{i=1}^N [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)]^2 \quad (2)$$

La minimizzazione si può effettuare con il calcolo differenziale.

Per vedere se la retta di regressione ben si adatta al modello si valutano dei parametri di bontà, ad esempio:

- **R^2** : misura la frazione di varianza campionaria di Y spiegata da X . È adimensionato e può variare tra 0 e 1. Se è pari a 1 significa che c'è un perfetto adattamento ai dati mentre se tende a zero vi è un basso adattamento. Affinché il modello di regressione sia abbastanza buono, questo parametro deve essere maggiore di 0,4.
- **SER** : è l'errore standard della regressione e misura la dispersione della distribuzione del residuo ε ovvero può essere inteso come la deviazione standard campionaria dei residui OLS.

E da notare però che la stima OLS si calcola su un campione di dati e se si considerasse un campione differente si avrebbe un risultato diverso dei parametri; ciò crea un'incertezza campionaria. Risulta dunque opportuno effettuare un *test sulla significatività* del campione trovato ovvero verificare l'ipotesi nulla che β_1 sia uguale a zero contro l'ipotesi alternativa che β_1 sia diverso da zero. Se l'ipotesi nulla non può essere rifiutata si esclude dunque un impatto causale della variabile X sulla variabile Y mentre in caso contrario si avrà un impatto con un'entità che dipende dal valore di β_1 .

La verifica dell'ipotesi può essere svolta in tre modi diversi, grazie al software *Stata*, e implica l'accettazione di una probabilità d'errore:

- In base al *p-value* si accetta o rifiuta l'ipotesi. Questo parametro indica la probabilità di rifiutare erroneamente l'ipotesi nulla quando invece è vera
- Alternativamente definito un livello di significatività α (pari a 1%,5% o 10%) di rifiutare in modo errato l'ipotesi nulla, si costruisce la *statistica t*, definita come il rapporto tra il coefficiente stimato e il suo errore standard

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} \quad (3)$$

Se confrontando il valore t ottenuto con quello della distribuzione t di student associato al livello di significatività α con $n-1$ gradi di libertà si ottiene che $|t| > t_{\alpha/2, n-1}$ allora si può rifiutare l'ipotesi nulla.

- In maniera analoga si può costruire l'intervallo di confidenza $(1 - \alpha)\%$, cioè un intervallo contenente il valore vero del coefficiente nel $(1 - \alpha)\%$ di tutti i possibili campioni estratti in modo casuale

$$I_{(1-\alpha)\%} = \hat{\beta}_1 \pm t_{\alpha/2, n-1} SE(\hat{\beta}_1) \quad (4)$$

Il metodo OLS permette di ottenere stimatori corretti dei parametri della popolazione solo se vengono soddisfatte le seguenti assunzioni:

1. La distribuzione di ε condizionata per ogni valore di X deve avere media nulla, ovvero la variabile X e l'errore ε non devono essere correlate tra loro. In altre parole, ε non deve comprendere fattori omessi, non considerati nel modello, correlati a X e che impattano Y .
2. Le coppie (X_i, Y_i) risultano indipendenti e identicamente distribuite, ottenute cioè da un campionamento casuale e appartenenti alla stessa popolazione.
3. Gli *outlier* in X e Y hanno momenti quarti finiti, cioè sono rari

Avendo a che fare con osservazioni reali e non con un esperimento ideale, è probabile che vi siano fattori omessi correlati con il regressore X e che vanno a determinare la variabile dipendente Y . Viene meno dunque la prima ipotesi dei minimi quadrati e conseguentemente lo stimatore OLS subirà la cosiddetta *distorsione da variabile omessa*. Per ovviare a ciò, è bene adottare un modello di regressione multipla che include variabili addizionali come regressori. Questo modello di regressione multipla rende possibile stimare l'effetto su Y della variazione di X mantenendo costanti gli altri regressori che influenzano Y e sono correlati con X .

Al fine di controllare quei fattori non osservati e costanti per caratteristiche idiosincratice di un certo gruppo di elementi all'interno del campione si inseriscono degli effetti fissi. Questi

vengono inseriti per sopperire alla mancanza di tutte quelle variabili omesse che sono differenti da un'entità all'altra ma rimangono costanti nel tempo. Analogamente aggiungendo gli effetti temporali sono controllate quelle variabili costanti tra le entità ma che cambiano nel tempo.

In caso di regressione multipla, il modello OLS mantiene le stesse assunzioni prima esposte con l'aggiunta che non ci sia collinearità perfetta tra le variabili considerate nel modello, ovvero che un regressore non sia funzione lineare esatta degli altri.

Si precisa che in questo studio, i calcoli tengono in considerazione un errore calcolato con una formula robusta all'eteroschedasticità, ossia risultano validi anche se la varianza dell'errore ε non è costante ma dipende dal valore assunto da X .

3.2 SPECIFICHE DEI MODELLI UTILIZZATI

Il modello OLS precedentemente analizzato ha l'obiettivo di fornire una stima dell'impatto causale dell'introduzione della fibra sull'occupazione in Italia.

Sono state presi in considerazione quattro diversi modelli. Ogni modello, rispetto a quello precedente, utilizza ulteriori fattori prima non presi in considerazione per valutare con maggiore accuratezza se l'introduzione della fibra abbia avuto un impatto significativo o meno. Si stima che inserendo effetti fissi sul comune, sul tempo e inserendo ulteriori controlli, la diffusione della fibra sia avvenuta in maniera casuale e che quindi i coefficienti stimati rappresentino effettivamente l'impatto causale della fibra sull'occupazione.

Il **primo modello** considerato è il seguente:

$$Y_{i,t} = \beta_0 + \beta_1 dummy_UBB_{i,t} + FE_t + FE_i + \varepsilon_{i,t} \quad (5)$$

con i che identifica il generico comune e t il generico anno compreso tra il 2019 e il 2021.

Le variabili considerate sono:

- $dummy_UBB_{i,t}$: è la variabile binaria di interesse che è pari a 1 se la fibra è disponibile nel comune i al tempo t e pari a 0 altrimenti.
- $Y_{i,t}$: è la variabile dipendente in forma logaritmica cioè il logaritmo degli occupati per ogni comune i nell'anno t .
- FE_t : controlla il trend di tipo temporale creando variabili binarie che identificano il singolo anno dal 2019 al 2021. Queste realizzano un controllo dei fattori dovuti al ciclo economico e suppliscono per tutte quelle variabili non osservabili che mutano nel tempo ma in maniera uniforme in tutti i comuni.
- FE_i : rappresenta l'effetto fisso sul comune. Viene creato un vettore di variabili binarie grande quanto il numero totale di comuni, in modo che ogni osservazione assumerà il valore 1 nella variabile corrispondente al suo comune e il valore 0 in tutte le altre. Questo effetto fisso controlla i fattori idiosincratici del comune evitando che vengano inserite variabili che influenzano la Y ma invariante nel tempo in quanto caratteristiche intrinseche del comune come ad esempio la posizione geografica.
- $\varepsilon_{i,t}$: costituisce il termine d'errore.

Avendo scelto la trasformazione logaritmica della variabile dipendente, β_1 costituirà la variazione “percentuale” della variabile dipendente a seguito dell'introduzione della fibra.

Nel **secondo modello** considerato, invece di considerare l'effetto fisso temporale si utilizza un effetto fisso dato dalla combinazione di anno e regione. Così facendo, si stima che i

cambiamenti che avvengono negli anni siano eterogenei per le diverse regioni italiane. Il modello in questione è il seguente:

$$Y_{i,t} = \beta_0 + \beta_1 dummy_UBB_{i,t} + FE_{t,regione} + FE_i + \varepsilon_{i,t} \quad (6)$$

Il **terzo modello** conserva il controllo temporale basato su regione e anno ma effettua un cambio della variabile di interesse. Invece di considerare l'impatto generato dalla disponibilità della banda ultralarga nel comune d'interesse considera l'effetto sulla variabile dipendente dovuto alla presenza di un anno in più di fibra. Viene dunque creata una nuova variabile indipendente *Years_UBB* che misura da quanti anni è disponibile la fibra nel comune. In sostanza, assumerà il valore 0 se nel comune non è ancora presente la fibra; 1 se la fibra diventa disponibile in quell'anno; 2 se il comune ha ricevuto la fibra l'anno precedente e così via.

$$Y_{i,t} = \beta_0 + \beta_1 Years_UBB_{i,t} + FE_{t,regione} + FE_i + \varepsilon_{i,t} \quad (7)$$

Nel **quarto modello** si considera inoltre il logaritmo della superficie del comune che interagisce con le dummy temporali. In tal modo si tiene in considerazione la diversa dinamica dell'occupazione in comuni con superfici diverse. Questo controllo è rilevante perché considera il fatto che i comuni che devono ancora ricevere la fibra, circa un quarto a fine 2021, sono generalmente quelli più piccoli del campione preso in esame.

$$Y_{i,t} = \beta_0 + \beta_1 Years_UBB_{i,t} + \beta_2 S'_{i,t} + FE_{t,regione} + FE_i + \varepsilon_{i,t} \quad (8)$$

Dopo un'analisi sulla totalità delle osservazioni, considerando che i risultati potrebbero essere "falsati" a causa di un'eccessiva aggregazione dei dati, si è deciso di verificare se i risultati differiscono in base all'area geografica in cui i comuni sono situati cioè Nord-Ovest, Nord-Est, Centro e Sud Italia.

Per ogni area geografica, viene rilevato l'effetto della banda ultralarga sull'occupazione nei comuni situati in queste aree utilizzando gli ultimi due modelli precedentemente descritti.

Infine, per effettuare un *test di robustezza*, il terzo e il quarto modello saranno poi validati utilizzandone uno analogo a *variabili strumentali* che utilizza come *strumento* la distanza tra il comune considerato e il più vicino nodo OPB (*Optical Packet Backbone*). Quest'ultimo modello verrà illustrato nel seguito di questo elaborato.

3.3 IL METODO TWO STAGE LEAST SQUARES

Affinché il coefficiente stimato dal metodo OLS generi una stima non distorta dell'effetto che la variabile d'interesse produce su quella dipendente è bene che non vi siano variabili omesse nel modello di regressione ovvero altri fattori non considerati che influenzano sia la variabile dipendente che quella esplicativa. Si è cercato di ovviare a questo problema di endogeneità da variabile omessa facendo dapprima ricorso ad una stima ad effetti fissi ed in seguito ricorrendo a trend temporali e a ulteriori controlli. Ci potrebbero essere però altre variabili omesse, ad esempio shock esterni, tipici del comune analizzato, che variano nel tempo e vanno ad influenzare sia la diffusione della fibra che il mercato del lavoro. Per esempio, la costruzione di un'industria in un certo comune produce indubbiamente un effetto positivo sul mercato del lavoro e può influenzare gli operatori di telecomunicazione ad investire nella banda ultralarga in queste zone. Una maggiore diffusione del contagio genera invece un impatto negativo sull'occupazione e una minore volontà degli operatori ad investire in fibra in queste aree. Questi eventi o altri simili, pur essendo specifici del comune considerato, non vengono rilevati dall'effetto fisso in quanto non sono costanti nel tempo e di conseguenza vanno a fare parte

dell'errore econometrico. Se tuttavia queste variabili non vengono considerate, la stima sarà distorta in quanto cattura l'effetto generato sulla variabile dipendente sia dalla variabile indipendente che da questi altri fattori.

Per ovviare a questo problema, si utilizza un approccio differente che mette la variabile non osservata nel termine d'errore, ma, anziché utilizzare il metodo OLS ne utilizza un altro che riconosce la presenza di variabili omesse; questo metodo è noto come *Two Stages Least Squares (TSLS)* o a *variabili strumentali (IV)*.

Di seguito verrà svolta un'analisi con questo metodo di stima. L'obiettivo è utilizzare questo metodo come test di robustezza per controllare se i risultati trovati tramite il metodo OLS vengono confermati.

3.2.1 DESCRIZIONE DEL METODO

Il metodo a variabili strumentali permette di ottenere uno stimatore consistente dei coefficienti ignoti della funzione di regressione nel caso in cui il regressore X sia correlato con l'errore ε . Per capire il funzionamento di tale metodo, bisogna considerare la variazione in X costituita da due parti una correlata all'errore ε e l'altra non correlata con questo. L'idea è quella di isolare l'informazione sulla variazione in X non correlata con ε , concentrandosi su di essa e trascurando quindi le variazioni in X che rendono distorta la stima tramite il metodo OLS. Ciò è reso possibile grazie all'adozione di una variabile *strumentale*. Quindi si può affermare che il metodo Two Stages Least Squares stima l'effetto generato da una variabile di interesse *endogena* utilizzando una variabile strumentale *esogena*, non correlata con l'errore, detta appunto *strumento*.

Affinché lo strumento sia valido, questo deve essere:

- *Rilevante*, cioè correlato con la variabile d'interesse: $\text{corr}(Z_i, X_i) \neq 0$
- *Esogeno*, cioè non correlato con l'errore econometrico, cioè con qualsiasi altro determinante della variabile dipendente Y : $\text{corr}(Z_i, u_i) = 0$

Se lo strumento soddisfa queste due condizioni si può utilizzare il metodo IV o dei minimi quadrati a due stadi per stimare il coefficiente della regressione.

Nel primo dei due stadi della regressione Two Stages Least Squares viene effettuata la regressione della variabile endogena X sullo strumento Z :

$$X_i = \pi_0 + \pi_1 Z_i + v_i \quad (9)$$

Questa regressione permette di capire se lo strumento è correlato con la variabile X ed inoltre restituisce la scomposizione richiesta di X . Viene infatti predetto un valore della variabile d'interesse \hat{X}_i pari a $\pi_0 + \pi_1 Z_i$ che non dipende da nessun'altra potenziale variabile omessa: \hat{X}_i non è quindi correlata con ε siccome Z_i è esogeno.

Nel secondo stadio viene eseguita la regressione tra il risultato predetto dal primo stadio e la variabile Y

$$Y_i = \beta_0 + \hat{\beta}_1 \hat{X}_i + \varepsilon_i \quad (10)$$

La differenza con l'equazione 1 consiste nel fatto che viene utilizzato \hat{X} invece di X e dunque si otterrà un coefficiente $\hat{\beta}_1$ che, escludendo tutti gli altri effetti endogeni, stima l'effetto reale del regressore X sulla variabile dipendente Y .

Un modo efficace per controllare la validità dello strumento consiste nel calcolare la statistica F al primo stadio della regressione. Questa statistica esegue un *test di ipotesi* per controllare se i coefficienti del modello sono entrambi nulli. Come regola empirica si considera forte lo strumento che presenta valori della statistica F superiori a 10.

3.2.2 SPECIFICAZIONE DEL MODELLO

Per effettuare una verifica dei risultati ottenuti tramite OLS e per risolvere la potenziale endogeneità della diffusione della banda ultralarga è stato implementato un metodo a variabili strumentali. La scelta dello strumento è stata orientata da diversi studi provenienti dalla letteratura sull'argomento disponibili nello studio *Ultra-fast broadband investment and adoption: A survey* di Abrardi e Cambini (2021). Come detto in precedenza, si è deciso di utilizzare la distanza fisica tra ogni comune e il più vicino nodo di telecomunicazione o *Optical Packet Backbone*. I nodi OPB sono localizzati secondo la precedente infrastruttura di telecomunicazione e la loro ubicazione è stata finalizzata nel 2012.

Per ciò che riguarda il modello di regressione adottato, sono state utilizzate due diverse specificazioni. La prima coincide con l'equazione 7 riguardante la regressione OLS. Ci si pone quindi l'obiettivo di trovare l'effetto che un anno in più di internet ultraveloce genera sull'occupazione considerando gli effetti fissi a livello di comune e i trend temporali a livello di regione.

$$Y_{i,t} = \beta_0^{\text{TSL}} + \beta_1^{\text{TSL}} \text{Years_UBB}_{i,t} + FE_{t,\text{regione}} + FE_i + \varepsilon_{i,t} \quad (11)$$

La seconda specificazione, coincidente con l'equazione 8 del modello OLS, considera un ulteriore controllo cioè il logaritmo della superficie del comune che interagisce con le dummy temporali. Si tiene così in considerazione il fatto che nei comuni con superficie differente si hanno trend diversi della diffusione della banda ultralarga e dell'occupazione.

$$Y_{i,t} = \beta_0^{\text{TSLS}} + \beta_1^{\text{TSLS}} \text{Years_UBB}_{i,t} + \beta_2^{\text{TSLS}} S'_{i,t} + FE_{t,\text{regione}} + FE_i + \varepsilon_{i,t} \quad (12)$$

Anche con questo metodo, al fine di ottenere risultati più approfonditi ed evitare una eccessiva aggregazione dei dati, sono stati analizzati gli effetti sia a livello nazionale che a livello di area geografica italiana.

3.2.3 ANALISI DELLO STRUMENTO UTILIZZATO

Per effettuare le regressioni esposte nel paragrafo precedente è necessario capire se lo strumento utilizzato, pari alla distanza tra il comune considerato e il più vicino nodo OPB, sia valido e in grado di spiegare una buona parte della varianza della variabile endogena. Ciò è possibile visualizzando per ciascuna suddivisione effettuata l'output del primo stadio della regressione. Si devono verificare le seguenti ipotesi:

- La variabile strumentale abbia un coefficiente statisticamente significativo. Si considera il fatto che la correlazione con la variabile *Years_UBB* abbia segno negativo in quanto per realizzare l'infrastruttura in fibra si rendono necessari lavori nel sottosuolo dunque risulta più conveniente coprire con la nuova tecnologia prima i comuni più prossimi al nodo OPB e solo successivamente quelli a distanza maggiore.

- La statistica F, che verifica se il coefficiente della variabile strumentale sia uguale a 0 nel primo stadio della regressione, non abbia un valore inferiore a 10. Ciò indicherebbe una debolezza dello strumento utilizzato.

Come ci si aspettava, dalle Tabelle 3.1 e 3.2 si evince che per tutte le suddivisioni effettuate applicando il metodo a variabili strumentali sia con controllo sulla superficie che senza la correlazione tra lo strumento Z e la variabile $Years_UBB$, che identifica il numero di anni di disponibilità della fibra, risulta significativa e negativa. Per tutte le regressioni effettuate la significatività mostra un p-value inferiore a 0.01. Il valore del test F per tutti i modelli di regressione risulta essere abbondantemente superiore a 10. Si può quindi affermare che lo strumento utilizzato risulta valido. Da notare come il valore del test F sia maggiore quando si effettua il controllo sulla superficie come mostrato in tabella 3.2 e conseguentemente i risultati ottenuti applicando l'equazione 12 risulteranno essere più precisi.

TUTTI I COMUNI ITALIANI	
	Years_UBB
Z	-0,00768*** (0,000246)
F-Test	227,31
Osservazioni	23427
Numero di comuni	7809

COMUNI DEL NORD OVEST	
	Years_UBB
Z	-0,01417*** (0,000613)
F-Test	239,68
Osservazioni	8907
Numero di comuni	2969

COMUNI DEL CENTRO	
	Years_UBB
Z	-0,00466*** (0,000712)
F-Test	157,18
Osservazioni	2853
Numero di comuni	951

COMUNI DEL NORD EST	
	Years_UBB
Z	-0,00889*** (0,000741)
F-Test	158,19
Osservazioni	4017
Numero di comuni	1339

COMUNI DEL SUD	
	Years_UBB
Z	-0,00619*** (0,000305)
F-Test	198,53
Osservazioni	7650
Numero di comuni	2550

Tabella 3.1: Correlazione tra lo strumento Z e la variabile di interesse Years_UBB – Metodo IV

(*** p<0,01, ** p<0,05, * p<0,1)

TUTTI I COMUNI ITALIANI	
	Years_UBB
Z	-0,00843*** (0,000243)
F-Test	231,73
Osservazioni	23427
Numero di comuni	7809

COMUNI DEL NORD OVEST	
	Years_UBB
Z	-0,01600*** (0,000594)
F-Test	285,67
Osservazioni	8907
Numero di comuni	2969

COMUNI DEL CENTRO	
	Years_UBB
Z	-0,00568*** (0,000712)
F-Test	173,95
Osservazioni	2853
Numero di comuni	951

COMUNI DEL NORD EST	
	Years_UBB
Z	-0,00917*** (0,000720)
F-Test	197,42
Osservazioni	4017
Numero di comuni	1339

COMUNI DEL SUD	
	Years_UBB
Z	-0,00680*** (0,000304)
F-Test	211,92
Osservazioni	7650
Numero di comuni	2550

Tabella 3.2: Correlazione tra lo strumento Z e la variabile di interesse Years_UBB – Metodo IV con controllo sulla superficie

(*** p<0,01, ** p<0,05, * p<0,1)

3.2 ANALISI DEI RISULTATI

In questo paragrafo vengono analizzati i risultati ottenuti dai modelli econometrici utilizzati, elencati nei precedenti paragrafi. Applicando il primo modello che considera come variabile d'interesse *dummy_ubb*, ovvero la presenza o l'assenza della banda ultralarga nel comune considerato, notiamo che a livello nazionale si registra un piccolo aumento dell'occupazione comunale pari allo 0,018%. A causa però dell'eccessiva aggregazione dei dati, tale aumento risulta caratterizzato da un basso grado di significatività con un valore del t di student pari a 0,44. Se invece di considerare l'effetto fisso del tempo, si considera l'effetto fisso dato dalla combinazione di anno e regione, ovvero il secondo modello preso in esame, notiamo che la stima diventa significativa ma negativa con un calo dell'occupazione pari allo 0,2%. Questo calo è spiegato dal fatto che la diffusione della banda ultralarga non è avvenuta in maniera omogenea nelle diverse regioni d'Italia come spiegato nel precedente capitolo. Un'altra motivazione di questo calo riguarda l'aumento di copertura della banda ultralarga nel periodo 2019-2020 a cui è però corrisposto, per via della pandemia, una diminuzione generalizzata nel numero di occupati che ha influenzato negativamente la stima tramite la variabile d'interesse *dummy_ubb*. In secondo luogo, bisogna considerare che i benefici sul mercato del lavoro derivanti dal dispiegamento di una nuova tecnologia, quale quella dell'Internet ultraveloce, non vengono riscontrati nell'immediato ma richiedono più anni. Nel breve periodo quindi si può rilevare che l'adozione di una connessione Internet veloce fa sì che vi siano modelli di lavoro meno *labour intensive* che adoperano quindi un minor numero di addetti. Inoltre, si può riscontrare un abbassamento delle barriere di uscita dal mercato con la conseguente fuoriuscita di aziende meno avanzate dal punto di vista tecnologico che non sono quindi state in grado di

cogliere i vantaggi derivanti dalla fibra ottica. Una panoramica dei risultati a livello nazionale finora esposti è raffigurata nelle tabelle 3.3 e 3.4.

TUTTI I COMUNI (effetti fissi su comune e anno)	
	Log(Occupati)
dummy_UBB	0,00018 (0,000411)
R-squared	0,35
Osservazioni	23427
Numero di comuni	7809

Tabella 3.3: Stime dei coefficienti ottenuti tramite il metodo OLS corrispondente all'equazione 5 con riferimento alla totalità dei comuni italiani

TUTTI I COMUNI (effetti fissi su comune e con interazione anno e regione)	
	Log(Occupati)
dummy_UBB	-0,002044*** (0,000409)
R-squared	0,51
Osservazioni	23427
Numero di comuni	7809

Tabella 3.4: Stime dei coefficienti ottenuti tramite il metodo OLS corrispondente all'equazione 6 con riferimento alla totalità dei comuni italiani

(*** p<0,01, ** p<0,05, * p<0,1)

Considerando invece come variabile d'interesse *Years_UBB* che considera il numero di anni in cui è presente la banda ultralarga in un dato comune, è possibile notare dalla Tabella 3.5 come sia a livello nazionale sia a livello di area geografica italiana vi sia una correlazione positiva con il logaritmo del numero di occupati per comune anche considerando gli effetti fissi dati dalla combinazione di anno e regione. È possibile dunque affermare che i benefici dovuti alla presenza di una connessione Internet ultraveloce non sono riscontrabili immediatamente tramite il passaggio da 0 a 1 della variabile indipendente *dummy_ubb* ma si registrano nel corso degli anni quindi sono maggiormente messi in luce tramite la variabile *Years_UBB*. Questa variabile considera maggiormente l'impatto sull'occupazione dei comuni che hanno ricevuto la fibra

prima del 2019, anno d'inizio dell'analisi svolta in questo lavoro. Si nota quindi che con l'aumentare del numero di anni di presenza dell'Internet veloce, questa tecnologia passa dall'essere un sostituto del lavoro ad esserne un complemento determinando quindi un aumento nel numero di occupati.

Come si può vedere dalla tabella sottostante i modelli OLS evidenziano correlazioni significative e positive della variabile *Years_UBB* sul logaritmo degli occupati per comune italiano. Utilizzando il più accurato modello OLS con controllo sulla superficie si vede come l'aumento dello 0,12% degli occupati per comune italiano dovuto alla diffusione della banda ultralarga sia dovuto principalmente alle regioni del Nord Ovest e del Sud che vedono un aumento percentuale dell'occupazione comunale dello 0,16% e dello 0,14% rispettivamente. Va fatto notare come proprio in queste aree si riscontri il maggior numero di occupati nel periodo 2019-2021. Le stime prodotte dal modello OLS vengono confermate dai 2 modelli a variabili strumentali corrispondenti alle equazioni 11 e 12. Questi due modelli, isolando il contributo della banda ultralarga, producono stime maggiori rispetto a quelle predette dal metodo dei minimi quadrati ordinari. Considerando tutti i comuni italiani, si nota un aumento dell'occupazione a livello comunale dovuto alla variabile *Years_UBB* dell'1% prendendo in considerazione il modello IV con controllo sulla superficie dei comuni. Con quest'ultimo modello si riscontra come l'aumento dell'occupazione comunale a livello nazionale sia questa volta dovuto maggiormente ai comuni situati nelle regioni del Centro e del Nord Ovest che vedono un aumento degli occupati del 2,91% e dell'1,26% rispettivamente. Isolando dunque l'impatto della banda ultralarga con il modello IV, attraverso l'utilizzo di una variabile strumentale non correlata con l'errore econometrico, si nota come nelle aree in cui la diffusione della fibra è maggiore come al Centro Italia l'effetto sull'occupazione risulta essere più evidente.

TUTTI I COMUNI ITALIANI				
	Log(Occupati) OLS	Log(Occupati) OLS con controllo	Log(Occupati) IV	Log(Occupati) IV con controllo
Years_UBB	0,00127*** (0,000225)	0,00125*** (0,000225)	0,01107*** (0,001387)	0,01003*** (0,001276)
Osservazioni	23427			
Numero di comuni	7809			

COMUNI DEL NORD OVEST				
	Log(Occupati) OLS	Log(Occupati) OLS con controllo	Log(Occupati) IV	Log(Occupati) IV con controllo
Years_UBB	0,00170*** (0,000383)	0,00169*** (0,000383)	0,01347*** (0,002077)	0,01264*** (0,000613)
Osservazioni	8907			
Numero di comuni	2969			

COMUNI DEL NORD EST				
	Log(Occupati) OLS	Log(Occupati) OLS con controllo	Log(Occupati) IV	Log(Occupati) IV con controllo
Years_UBB	0,00060* (0,000410)	0,00052* (0,000410)	0,01007** (0,003082)	0,00694** (0,000741)
Osservazioni	4017			
Numero di comuni	1339			

COMUNI DEL CENTRO				
	Log(Occupati) OLS	Log(Occupati) OLS con controllo	Log(Occupati) IV	Log(Occupati) IV con controllo
Years_UBB	0,00080* (0,000520)	0,00075* (0,000520)	0,03598** (0,012337)	0,02915*** (0,008699)
Osservazioni	2853			
Numero di comuni	951			

COMUNI DEL SUD				
	Log(Occupati) OLS	Log(Occupati) OLS con controllo	Log(Occupati) IV	Log(Occupati) IV con controllo
Years_UBB	0,00149*** (0,000414)	0,00148*** (0,000414)	0,00739*** (0,002061)	0,00700*** (0,001878)
Osservazioni	7650			
Numero di comuni	2550			

Tabella 3.5: Stime dei coefficienti ottenuti tramite il metodo OLS e il metodo IV corrispondenti alle equazioni 7,8,11,12

(*** p<0,01, ** p<0,05, * p<0,1)

CONCLUSIONI

L'obiettivo di questo lavoro è stato quello di verificare, se esiste, una correlazione tra l'accesso ad una connessione a banda ultralarga e il livello di occupazione per comune italiano e di quantificarne l'entità.

Nella prima parte, è stata effettuata un'analisi dei risultati emersi da altri studi riguardanti l'impatto della disponibilità di Internet sul grado di occupazione. Sono state quindi descritte le politiche nazionali ed europee atte a promuovere lo sviluppo della connessione in fibra ed è stata effettuata una panoramica sullo sviluppo di tale tecnologia in Italia.

L'analisi è stata effettuata su un dataset comprendente 7809 comuni italiani di cui si avevano a disposizione i dati di copertura della banda ultralarga provenienti dal sito di *Infratel* e quelli sul numero di occupati disponibili sul sito del *Ministero dell'Economia e delle Finanze (MEF)*. Entrambi i dataset fanno riferimento al triennio che va dal 2019 al 2021. Questi dati sono quindi stati analizzati approfonditamente rilevando trend temporali sia a livello nazionale che a livello di area geografica italiana.

L'analisi econometrica è stata effettuata utilizzando due diversi metodi di stima. Il primo metodo è stato l'*Ordinary Least Squares (OLS)* dopo di che per confermare i risultati ottenuti sono state effettuate le stesse regressioni ma con il metodo *Two Stages Least Squares (TSLS)*. Come strumento per svolgere le analisi a variabili strumentali è stata scelta la distanza tra un generico comune e l'*Optical Packet Backbone (OPB)* più vicino. Al fine di cogliere gli effetti sull'occupazione con un livello di dettaglio via via maggiore sono state implementate diverse specifiche del modello utilizzato. In particolare, si sono considerati effetti fissi a livello di comune, trend temporali per regione ed anche trend differenziali per superficie comunale.

I risultati ottenuti dimostrano come i benefici sull'occupazione dovuti alla disponibilità di una connessione a banda ultralarga non siano riscontrabili nell'immediato e risentono dell'eterogeneità della diffusione della banda ultralarga nelle diverse regioni italiane. Questi effetti si riscontrano nel corso degli anni e vengono messi in luce utilizzando la variabile indipendente *Years_UBB* che considera gli anni di disponibilità di una connessione in fibra in un dato comune. I benefici sull'occupazione registrati utilizzando questa variabile con il modello OLS sono stati confermati da quello a variabili strumentali. In particolare, con il modello OLS l'impatto maggiore sull'occupazione è stato nei comuni situati al Nord Ovest e al Sud dove il numero di occupati è risultato più alto rispetto alle altre aree nel triennio 2019-2021. Con il metodo *TSLS*, isolando la quota della variabile indipendente non correlata con l'errore econometrico, si ottengono stime maggiori rispetto al modello OLS e con benefici sull'occupazione più rilevanti nei comuni situati al Centro Italia dove la diffusione della banda ultralarga è stata maggiore nei tre anni considerati.

A fronte di questi risultati, si può dunque affermare che gli investimenti in connessioni ultraveloci producono effetti positivi sul piano occupazionale e permettono ai comuni di essere più resilienti a seguito di shock esterni come quello pandemico recentemente avvenuto. Queste infrastrutture costituiscono dunque un asset fondamentale e strategico per il mercato del lavoro nazionale.

Bibliografia

- Abrardi, L., & Cambini, C. (2019). Ultra-fast broadband investment and adoption: A survey. *Telecommunications Policy*, 43(3), 183-198.
- Camera dei deputati (2021). La banda ultralarga, il 5G e le reti di comunicazione elettronica. <https://temi.camera.it/>
- Ciappanna, E., & Colonna, F. (2020). Connettere l'Italia. *Rivista di politica economica*, (1), 93-105.
- Corte dei conti Europea (2018). La banda larga negli Stati membri dell'UE (12/2018)
- Istat (2021). *Annuario statistico italiano 2021*.
- Istat. Tasso di disoccupazione. Anni 2019-2021. <https://dati.istat.it/>
- Istat. Popolazione residente al 1° gennaio per fascia di età. Anni 2019-2021. <https://dati.istat.it/>
- MISE. Piano Strategico Banda Ultra Larga. Tecnologia. <https://bandaultralarga.italia.it/>
- MISE (2015). Il piano di investimento per la diffusione della banda Ultralarga.
- MITD & MISE (2021). Strategia italiana per la banda Ultralarga “Verso la Gigabit Society”.
- MITD (2021). Piano “Italia a 1 Giga”
- MITD. Banda Ultra Larga. La Strategia Italiana per la Banda Ultra Larga. <https://innovazione.gov.it/>
- TIM (2018). Netbook la rete per l'Italia.
- TIM (2021). La rete per l'Italia Netbook 2021.
- Bertschek, I., Briglauer, W., Hüschelrath, K., Kauf, B., & Niebel, T. (2015). The economic impacts of broadband internet: A survey. *Review of Network Economics*, 14(4), 201-227.

Czernich, N. (2014). Does broadband internet reduce the unemployment rate? Evidence for Germany. *Information Economics and Policy*, 29, 32-45.

Kolko, J. (2012). Broadband and local growth. *Journal of Urban Economics*, 71(1), 100-113.

Stock, J. H., & Watson, M. W. (2005). *Introduzione all'econometria*. Pearson Italia Spa.