Information Economics and Policy 68 (2024) 101094

Contents lists available at ScienceDirect

Information Economics and Policy

journal homepage: www.elsevier.com/locate/iep

Do Gigabits Mean Business? “Ultra-Fast” broadband availability’s effect on business births

Christina Biedny, Brian E. Whitacre *, Andrew J. Van Leuven

Department of Agricultural Economics, Oklahoma State University, USA

ARTICLE INFO

JEL Codes:
D22
O33
R3
R10

Keywords:
Digital divide
Broadband
Gigabit
Event study
Business births

ABSTRACT

Recent U.S. broadband programs prioritize high-speed infrastructure with download speeds of 100 megabits per second (Mbps) or more. Some internet providers have already built broadband networks capable of gigabit speeds (1000 Mbps) and more of this infrastructure is likely forthcoming due to increased federal support. The availability of this “ultra-fast” internet may have important implications for business creation. Business-level data from Data Axle is combined with the Federal Communication Commission’s Form 477 broadband availability data and aggregated to the census block group level to form a six-year panel from 2015 to 2020. We examine the effects of three thresholds of broadband speed availability on business births per 10,000 population in eight U.S. states using an event study design. Results suggest that 100+ Mbps availability increases business births for at least up to five years after being introduced. The impacts are largest in metro block groups and for select industries. Relative to block groups treated at the 100 Mbps level, access to 250 Mbps shows additional benefits for business creation; however, the results for gigabit speed provision are less conclusive. This may change as technology continues to evolve and ultra-fast speed becomes more necessary for business operations.

Introduction

New business creation is a vital component of economic growth (Sternberg and Wennekers, 2005; Acs, 2006; Urbano et al., 2019), and federal and local governments often seek policy levers that can emphasize entrepreneurship. Improving local broadband capability is one such lever, based on the idea that this technology can change and enhance business capabilities and provide individuals with tools necessary for growth – including new business opportunities. Research has shown that broadband internet access improves a number of economic outcomes including rural housing values (Deller and Whitacre, 2019; Wolf and Irwin, 2023), unemployment rates (Hasbi and Bohlin, 2022; Lobo et al., 2020; Czernich, 2014), firm location decisions (Kim and Orazem, 2016), firm productivity (Fabling and Grimes, 2021), jobs and income (Whitacre et al., 2014a, 2014b), employment growth (Rupasingha et al., 2023), and other metrics of regional development and firm performance (see Bertschek et al., 2015 for a survey, or Mack et al., 2023 for a review related to rural development topics).

Recent studies have examined the impacts of broadband on business creation, but generally lack an assessment of the speeds required for such change. In a study of 58 countries, Alderete (2017) found that mobile broadband access between 2007 and 2012 had a positive influence on entrepreneurial activity. Broadband also increased the establishment birth rate for women-led businesses and establishments in rural areas between 2003 and 2005 (Conroy and Low, 2022) and has even been touted as more important infrastructure than highways and railroads for increasing business start-ups (Audretsch et al., 2015). Deller et al. (2022) show that broadband speed availability has a positive impact on business creation in rural U.S. counties as of 2014 (when speeds were relatively limited), but their findings reinforce that these results vary across industries. These studies, though insightful, utilize data covering broadband speeds lower than recent U.S. policy requires. The issue is no longer one of strictly internet access, but of internet speed and connection quality.

The Federal Communication Commission’s (FCC) definition of “high speed internet” continues to be defined as 25 Megabits per second

Abbreviations: ACS, American community survey; BDC, Broadband data collection; CEM, Coarsened exact matching; DATA, Deployment accuracy and technological ability; FCC, Federal Communications Commission; GB, Gigabit; ICT, Information and Communications Technology; IIJA, Infrastructure Investment and Jobs Act; ISP, Internet service provider; Mbps, Megabits per second; NAICS, North American Industry Classification System; NETS, National establishment time series, PSM, Propensity score matching; QCEW, Quarterly Census of Employment and Wages; RCT, Randomized control trials; TWFE, Two-way fixed effects.

* Corresponding author at: 504 Ag Hall, OSU Campus, Stillwater, OK 74078, USA.
E-mail address: brian.whitacre@okstate.edu (B.E. Whitacre).

https://doi.org/10.1016/j.infoecopol.2024.101094
Received 17 August 2023; Received in revised form 12 March 2024; Accepted 30 May 2024
Available online 6 June 2024

(Mbps) download and 3 Mbps upload, or 25/3 (FCC, 2015).[1] This threshold is now widely viewed as insufficient in a society reliant on the internet for numerous devices running simultaneously.[2] As a result, the 2021 Infrastructure Investment and Jobs Act (IIJA) requires broadband projects that it funds to provide minimum download/upload speeds of 100/20 Mbps (IIJA, P.L. 117-58). Allocation amounts for these projects were announced in June 2023 for each of the more than 50 U.S. states and territories, totaling over $42 billion to be spent on broadband over the next five years (NTIA, 2023). While the legislation requires minimum download speeds of 100 Mbps for these projects, many will likely be built for significantly higher speeds (Whipple, 2023) – including those exceeding 1000 Mbps (1 Gigabit per second, or Gbps). Since the infrastructure needed for these speeds is costly and typically installed in densely populated urban areas where installation is more cost effective,[3] an important but unanswered question is whether prior investments in such “ultra-fast” broadband has resulted in new business creation.

This study combines U.S. business-level data from Data Axle with FCC broadband availability data to determine to what extent available internet speeds impact the number of business “births” (startups) from 2015 to 2020 in eight states in and adjacent to the “Southern Plains” region (see Fig. 1).[4] In line with previous studies (Duvivier et al., 2021; Deller et al., 2022), we consider separate specifications for the number of business births by industry since the effects of ultra-fast internet service may be heterogenous across industries. Equally important is the consideration of whether impacts vary for different speed thresholds. Though we are particularly interested in whether gigabit speeds affect business births, build-out and availability of gigabit service is still somewhat limited and recent U.S. legislation prioritizes access to 100 Mbps. Therefore, we consider other “fast” download speed thresholds including 250 Mbps and compare the results. We limit our analysis to geographies receiving broadband service after our time period of interest begins in order to assess the effects immediately after its introduction, as well as over the years that follow.

This paper contributes to the Information and Communication Technology (ICT)-related literature in several ways. First, existing U.S. studies that focus on broadband’s effect on business births typically use data from before gigabit internet speeds were possible. Several recent studies consider “ultra-fast” internet speeds, but these have investigated the relationship of broadband on economic outcomes outside of the U.S. (Duvivier and Bussière, 2022; Duvivier et al., 2021; Falk and Hagsten, 2021; Maude, 2020; Sarachuk and Missler-Behr, 2020; Galloway et al., 2011).[5] As such, we seek to fill a gap in the consideration of U.S. “ultra-fast” internet. Second, this research uses establishment-level data analyzed at the census block group level across a subset of eight states.[6] Previous U.S. broadband-related research has considered new business formation but mostly focuses on single states or higher-level geographies such as counties, which may mask more local effects. Third, we add to the literature specifically focused on broadband impacts in rural locales. One recent study found that lower broadband availability is a principal factor for lower innovation rates in rural counties compared to large cities in the U.S. (Han et al., 2023). Since the IIJA seeks to encourage build out specifically to underserved areas and these areas are typically those outside of major metropolitan areas, it is important to understand what non-metro areas may gain from broadband expansion. Studies focused on broadband and the rural economy are relatively few and have achieved mixed results.[7] Finally, results from this study may inform broadband policy decisions relevant to the current speed environment. Recent empirical evidence on business births and faster broadband speeds will be useful to local policymakers as they decide where and how their share of available broadband funds should be deployed.

Data

To estimate the impact of high-speed broadband on business births, we combine data from two sources: Data Axle’s U.S. Businesses dataset and the FCC’s Form 477 broadband availability reporting. The analysis is conducted at the census block group level, with the annual number of new businesses per 10,000 population serving as the outcome variable. Census block groups contain between 600 and 3000 people, and there are an average of 4795 block groups per state (U.S. Census Bureau 2022). The resulting panel dataset includes businesses in nearly 36,000 census block groups in eight states from 2015 to 2020.

Business-level data for select industries was purchased from Data Axle (formerly Infogroup), a data, technology, and marketing services provider (Data Axle, 2022). Data Axle’s U.S. Businesses dataset is compiled from telephone directories and other sources of business registration information including secretaries of state, county courthouse filings and public records. Data Axle’s verification teams are responsible for calling businesses to gather, verify and update information each year.[8] The dataset contains millions of U.S. businesses with information such as number of employees, year the business was established and North American Industry Classification (NAICS) industry code.[9] Data Axle business data has been used in the regional science literature on topics ranging from the impact of Main Street programs to commercial effects of natural disasters to understanding disparities in food access (Van Leuven, 2021; Meltzer et al., 2021; Dharmasankar and Yoo, 2023; Taylor et al., 2022; Theodos et al., 2022). Prior work has noted that this data is more expansive than other datasets such as County Business Patterns because it is more likely to capture sole proprietorships or small chain establishments (Meltzer et al., 2021). The nature of the data (detailed addresses; 6-digit NAICS codes, and year established) allows for tracking of new business births over time. The dataset purchased includes businesses categorized under eight different

[1] This official definition changed in March 2024 to 100 Mbps download and 20 Mbps upload (FCC, 2024).

[2] Deloitte (2022) reports that on average, U.S. households have an average of 22 connected devices. Further, the FCC Chairwoman proposed raising the official U.S. broadband threshold to 100/20 in 2022 but this has not been enacted to date (O’Shea, 2022).

[3] Per estimates from the U.S. Department of Transportation (DOT), the average cost of installing the fiberoptic cable necessary to deliver these speeds is $27,000 per mile (see U.S.DOT, 2022 for detailed data).

[4] An “establishment” or “business” represents a single physical location that offers goods or services. A business may be part of a multi-establishment firm, so we use the terms establishment and business interchangeably here to clarify that we are not concerned with firm-level start-up activity.

[5] The European Union and United Kingdom distinguish different categories of broadband speeds than the U.S. “Basic” broadband is any speeds less than 30 Mbps, “fast” broadband is between 30 – 100 Mbps, and “ultrafast” broadband is anything faster than 100 Mbps (Hasbi, 2020). As such, the new U.S. categorization of “high-speed” internet aligns with the term “ultra-fast” broadband in much of Europe.

[6] A census block group is the second smallest geographic area for which data is collected by the U.S. Census Bureau. A census block group is made up of census blocks. In the 2010 census, a block group was made up of 51 blocks on average. A block group is a subdivision of a census tract (the next higher-level geography). (U.S. Census Bureau, 2022)

[7] Availability of broadband increased rural economic growth in several studies (Stenberg 2009; Kolko, 2012; Atasoy, 2013; Lobo et al., 2020) while others found no effect (Kandilov and Renkow, 2010; Mahasuweerachai et al., 2010; Whitacre, 2011; Conley and Whitacre, 2020). Mack et al. (2023) review much of this literature.

[8] Authors’ discussion with Data Axle relationship manager, November 2022.

[9] NAICS uses a coding system to classify all economic activity into twenty, 2-digit industry sectors.

Fig. 1. Eight-state study area.

two-digit industry codes (Table 1). These industries were chosen due to the importance of the internet for business operations relative to other industries (Shideler and Badasyan, 2012; Deller et al., 2022).

The dependent variable of interest, business births per 10,000 population, was constructed by assessing the number of businesses in each block group that are newly added to the dataset since the prior year. Table 2 summarizes business births per 10,000 population across the block groups in our final sample during the 2015–2020 period as well as by specific industry.[10] Health Care & Social Services had the largest number of business births per capita while Information and Manufacturing had the fewest. It is worth noting that while the aggregate dataset did not have a high frequency of zeros (no business births in a block group since the prior year), some subsamples – especially by industry – do have a higher proportion of zeros. For example, the aggregate dataset had only 42 % of observed zeros in the dependent variable while manufacturing (91 %), and information (94 %) were considerably higher.

Table 1
North American Industrial Classification System (NAICS) industry codes included in Data Axle data.

Industry	2-digit NAICS industry code
Manufacturing	31–33
Retail Trade	44–45
Information	51
Finance & Insurance	52
Real Estate & Rental/Leasing	53
Professional, Scientific & Technical Services	54
Health Care & Social Services	62
Accommodation & Food Services	72

Source: U.S. Census Bureau.

Table 2
Summary statistics – Annual business births per 10,000 population (Block Group).

Annual business births per 10,000 population	Mean	Standard Deviation	Min	Max	% of 0s
Full Sample	0.0381	0.2143	0.00	19.01	41.79
By industry:					
Manufacturing	0.0013	0.0106	0.00	0.68	90.95
Retail Trade	0.0058	0.0305	0.00	2.21	71.68
Information	0.0010	0.0096	0.00	0.83	93.65
Finance & Insurance	0.0033	0.0372	0.00	6.62	85.03
Real Estate & Rental/ Leasing	0.0030	0.0210	0.00	1.50	86.96
Professional, Scientific & Technical Services	0.0051	0.0315	0.00	2.32	79.79
Health Care & Social Services	0.0155	0.1469	0.00	16.49	73.37
Accommodation & Food Services	0.0031	0.0196	0.00	1.21	83.68

Source: Data Axle.

Broadband availability data for the time period of interest are publicly available and sourced from the FCC's Form 477 (FCC, 2022a). Between 2014 and 2022, all broadband providers were required to file Form 477 with the FCC twice per year in June and December. The filing contains information on census blocks where an internet service provider (ISP) can or does offer service to *at least one location* within the block. This means that if a single location can be offered service, the entire census block is considered "served." Form 477 data has been widely criticized for this "one served, all served" approach since it may overstate actual internet availability (Busby and Tanberk, 2020; Busby et al., 2021; Major et al., 2020). Ford (2022) argues that this problem is more pronounced in rural areas, which have larger Census blocks. Form 477 also relies on ISPs reporting maximum *advertised* speeds rather than actual user-experienced speeds. At least 15 % of internet users nationwide may be receiving less than their internet provider's maximum advertised speeds, though a number of complex dynamics impact these

[10] These summary statistics represent only block groups left after dropping those with 100 Mbps service prior to 2015, as discussed in the data section.

conclusions (Supan, 2021; Gallardo and Whitacre, 2024).

Despite the shortcomings of the FCC's Form 477, it represents the only large-scale aggregation of U.S. broadband availability data to date. Reporting requirements have changed under recent legislation, however.[11] The new data is expected to overcome the "one served, all served" criticisms of current Form 477 data and is subject to a challenge process in which providers, local governments and other eligible entities can file challenges to either correct or supplement the provider data.[12] This new process was used to identify more than 11.8 million "locations" (i.e. business or residential buildings where fixed broadband is, or could be, installed) lacking 100/20 Mbps service that served as the basis for distributing the IIJA's $42.5 billion broadband-focused funding in June 2023 (Conlow, 2023; NTIA, 2023).

Our primary interest is in the impacts of three distinct "max advertised download speed" thresholds (provided by any wired technology): 100+ Mbps; 250 Mbps, and 1 Gigabit (1000 Mbps). We consider a block group to be "treated" if at least 50 % of establishments within the block group have access to maximum download speeds for each threshold.[13] In order to isolate the effect of higher speeds, the treated block groups must be compared only to those block groups that do not meet the threshold. The analysis thus focuses on three distinct comparison groups:

- Block groups with 100 Mbps or greater versus those block groups with less than 100 Mbps,
- Block groups with 250 Mbps (but not 1 GB) versus block groups with 100 Mbps, and
- Block groups with 1 Gbps versus those with 250 Mbps.

Importantly, we remove from the analysis any block group that had already received at least 100 Mbps service prior to 2015 (using 2014 Form 477 data). This removal of over 23,000 mostly urban block groups significantly reduces our sample size but allows us to focus on the introduction of "new" broadband service and the resulting potential impact on business births.

Table 3 displays the number of treated and control block groups for each speed threshold, by year. Note that the number of treated block groups increases over time in each category as infrastructure investment and broadband availability diffused during the period of study. Further, the control groups only include block groups that have met some threshold of lower available speeds (i.e., the gigabit control group only includes block groups with at least 250 Mbps) – and as such tend to decrease over time. The final groupings (as of 2020) are displayed visually in Fig. 2. The first treated category (100+ Mbps) represents *all* high-speed investment, so the 11,670 block groups are made up of the 1942 that received 100 Mbps, the 3529 that received between 250 Mbps and 1000 Mbps, and the 6199 that received over 1000 Mbps. The remaining treated groups (250 Mbps; 1000 Mbps) meet higher-speed thresholds only and limit the control group comparisons to block groups meeting specific lower-tier speeds.

Table 3
Number of Treated and Control block groups before matching.

	100+ Mbps (vs <100)		250 Mbps (vs. 100)		Gigabit (vs. 250)	
Year	Treated	Control	Treated	Control	Treated	Control
2015	3,298	10,369	307	5,164	411	9,317
2016	8,014	5,653	364	5,107	1,350	8,378
2017	8,954	4,713	1,586	3,885	2,691	7,037
2018	9,643	4,024	3,048	2,423	3,810	5,918
2019	10,798	2,896	3,162	2,309	5,174	4,554
2020	11,670	1,997	3,529	1,942	6,199	3,529

Source: Authors' calculations from Data Axle & FCC (2015–2020) data.

Census block group demographic covariates are summarized in Table 4. Covariates including block group population, race, percent of the population 25 and over with a bachelor's degree or higher, percent of households with a broadband connection, percent of the population in poverty, median household income, and percent of employed individuals age 16+ were sourced from the U.S. Census Bureau's five-year American Community Survey (ACS) for 2014–2018. Other covariates include baseline download speed in 2015 (from FCC Form 477), the total number of businesses, the proportion of businesses considered "goods producing" (LODES[14]), and rural-urban continuum code (USDA ERS).

Methods

Although randomized controlled trials (RCTs) are often considered the "gold standard" approach for estimating causal effects of treatments on outcomes, they are often expensive and/or difficult or impossible to conduct. With a treatment such as broadband availability, an RCT would be both expensive and difficult, if not impossible to conduct – especially across multiple states. Instead, we use observational data which requires quasi-experimental methods to arrive at potentially causal conclusions.

Panel event study models have become a popular quasi-experimental tool, evolving from the two-way fixed effects (TWFE, also known as difference-in-difference) model. The event study model is concerned with estimating the impact of an event which may occur at different times for different units.[15] For a treatment that may occur at different points in time for different units, the TWFE model can be written as:

$$y_{it} = \alpha + \beta \cdot Post_{it} + \mu_i + \tau_t + \boldsymbol{X}_{it}\boldsymbol{\delta} + \varepsilon_{it} \quad (1)$$

where y_{it} is the outcome of interest for unit i at time t, $Post_{it}$ serves as a binary indicator of "post-event" equal to one in all periods after the occurrence of the event in treated units,[16] μ_i and τ_t are unit and time fixed effects, respectively, $\boldsymbol{X}_{it}$ is a vector of optional controls, and ε_{it} is an unobserved error term.

The panel event study design expands on this TWFE model by including parameters on lead and lag time periods before and after the event of interest.[17] The inclusion of such leads and lags provides two advantages over the TWFE specification. First, the TWFE model relies heavily on the parallel trends assumption: that without treatment, differences in outcomes between treated and control groups would have

[11] The Broadband Data Collection (BDC) program, established under the Broadband Deployment Accuracy and Technological Availability (DATA) Act of 2020, requires all providers of fixed and mobile internet connections to file shapefiles and/or propagation maps and models that represent the service area of the provider (FCC, 2022b). The BDC program also required the development of an address-level "broadband serviceable location fabric" that provides more granularity than the widely criticized block-level aggregation (Whitacre and Biedny, 2022).

[12] The *bulk* challenge process was opened for the first time on September 12, 2022. Bulk challenges are meant to help expedite the review of challenges to multiple serviceable locations. Amended maps were published in November 2022 and a rolling challenge process allows eligible entities to file challenges for individual locations (FCC, 2022c).

[13] Larger percentage thresholds (75%, 90%) are tested as robustness checks.

[14] LODES or "LEHD Origin-Destination Employment Statistics" is a data product of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD).

[15] The canonical difference-in-difference model, popularized by Ashenfelter and Card (1985), considers an event that takes place at the same time for all treated units. The models discussed here are an expansion to include consideration of treatment that may occur at different points in time.

[16] $Post_{it} = 1[t \geq Event_i]$ where $Event_i$ is a variable recording the time period t in which the event occurs in/for unit i.

[17] Note that "leads" are defined as time periods *before* receipt of treatment while "lags" are time periods *after* receipt of treatment.

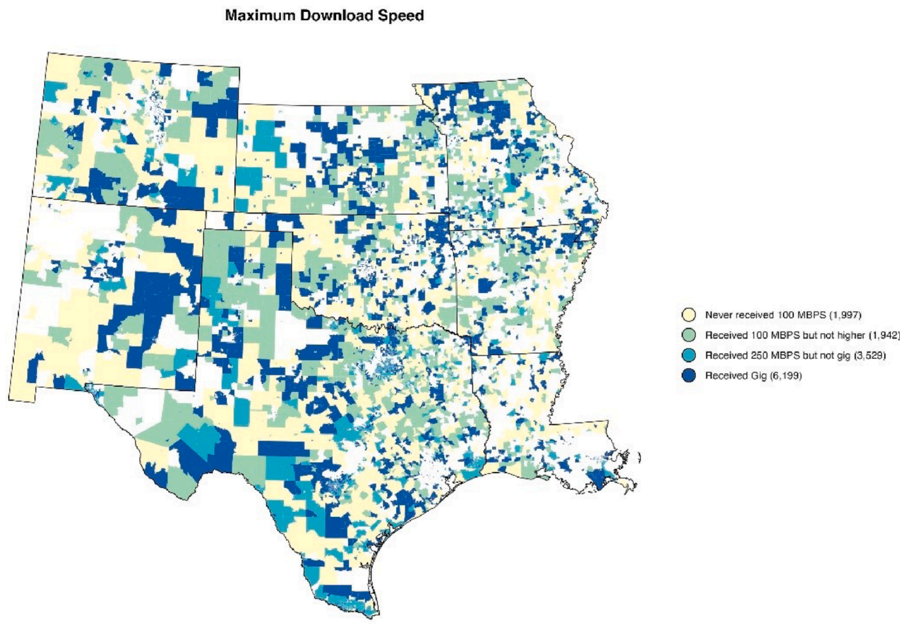

Fig. 2. Broadband speed availability across study area, by block group
Block groups with at least 100 Mbps prior to 2015 are omitted (white).

Table 4
Summary statistics – Block group demographic variables.

Variable	Mean	Std. Dev.	Min.	Max.
Population (2014–2018)	1,497	1,003	0	24,689
% non-white	0.4359	0.3241	0	1
% bachelor's degree or higher	0.2147	0.1636	0	1
% households with broadband connection	0.5015	0.2077	0	1
Total businesses	32.6955	82.2722	0	3,351
Goods-producing share of businesses	0.2326	0.2316	0	1
Baseline (2015) download speed (Mbps)	214	319	0	1,000
% in poverty	0.1785	0.1407	0	1
Median Household Income	$53,982	$26,502	$0	$250,001
% employed	0.5974	0.1220	0	1
Rural-urban Continuum Code (1–9)	3.4098	2.4529	1	9

Source: U.S. Census Bureau's American Community Survey 2014–2018; FCC Form 477; USDA Economic Research Service; U.S. Census Bureau's Longitudinal Employer-Household Dynamics: Employment Dynamics Origin Destination Employment Statistics (LODES).

remained constant over time. However, Eq. (1) does not allow for simple inspection of pre-treatment trends whereas a specification that includes distinct parameters for each period prior to treatment does. Second, the inclusion of lags allows us to consider the potentially dynamic effects after the treatment/event takes place. This may include whether an effect grows or shrinks over time after treatment and whether such effects are short-lived or more permanent. These dynamic outcomes are particularly important for policymakers, who are likely interested in knowing if the impacts of an intervention last only a short time.

Our data is comprised of a panel of census block groups b and time periods (years) t. We seek to estimate the impact of gigabit (or 250 Mbps, 100 Mbps) internet speed availability which may be realized at different times in different block groups. The event study specification can be written as:

$$births_{bt} = \alpha + \sum_{j=2}^{J} \beta_j \cdot Lead_{bt}^{j} + \sum_{k=1}^{K} \gamma_k \cdot Lag_{bt}^{k} + \mu_b + \tau_t + \varepsilon_{bt} \tag{2}$$

where $births_{bt}$ represents business births per 10,000 population in census block group b at year t, μ_b and τ_t are block group and time fixed effects, respectively, and ε_{bt} is an unobserved error term. $Lead_{bt}^{j}$ and Lag_{bt}^{k} are a set of binary variables that indicate that a block group is a given number of periods away from being treated with a specific broadband speed in a given time period. Leads denote periods prior to treatment and lags denote periods after treatment.[18] In Eq. (2), J leads and K lags are included. Since our panel covers the period 2015 – 2020, we consider up to $J = 4$ leads and $K = 5$ lags. Like other dummy variables, a single lead (or lag) must be omitted to capture the baseline difference between block groups where service of a given speed is and is not available. In Eq. (2), we omit the first lead where $j = 1$; the period immediately preceding treatment for each block group. Omitting this lead is conventional and allows for inspection of whether the treatment had an immediate impact (i.e., in time = 0). The β_j's and γ_k's are the coefficients to be estimated for each of the leads and lags to the receipt of the broadband threshold under consideration. Appendix A provides a stylized example of the lead and lag variables for hypothetical block groups A through D.

When estimating causal effects using observational data, it is important to control for characteristics within the treated portion of the sample that may systematically differ from the untreated portion. Failure to address such differences introduces a form of selection bias which, if left unaddressed, may result in misleading conclusions (Ford, 2018).[19] Matching provides a non-parametric way to implement such controls and create balance. Popular matching methods including Propensity Score Matching (PSM) and Mahalanobis matching can be used, however these are iterative procedures which only improve the average imbalance within the data. Alternatively, Coarsened Exact Matching (CEM) serves as a "monotonic imbalance-reducing matching method" which reduces the potential bias inflicted by imbalanced data (Iacus et al.,

[18] Recall that in this study, "receipt of broadband" refers to at least 50% of businesses within a block group having speeds greater than 100 Mbps/250 Mbps/1 Gigabit available in the census block in which they reside. Robustness checks using thresholds of 75% and 90% of businesses within a block group were also conducted (see Appendix B).

[19] Ford (2018) argues that the results of a previous study of broadband and employment by Crandall et al. (2007) were misleading due to the authors' failure to address such selection bias.

2009). CEM allows the user to choose specific demographic covariates on which the matched control groups are generated rather than the iterative "guess-and-check" process required of other matching methods.[20] This methodology allows the control groups to be selected so that they are similar to the treatment group in terms of demographic characteristics that may also play a role in business creation. In practice, each observation in the treated and control groups are assigned a "bin signature" representing its placement in the distribution of each covariate. Observations are removed from the CEM sample when a specific "bin signature" contains only one type (treated or control, i.e., a match cannot be made). In this way, only treated and control block groups that are similar in terms of the selected covariates remain in the matched sample. The covariates considered for matching in each model specification include block group population, race, percent of the population with a bachelor's degree or higher, percent of households with a broadband connection, the total number of businesses in the block group, and the proportion of block groups designated as "goods producing", among others.[21] As robustness checks, we also conduct CEM using sets of covariates that would be considered more and less restrictive than the baseline. Table 5 displays each of the matching covariates and the specification(s) to which they belong. Note that the more restrictive matching process results in the inclusion of slightly more than half as many total block groups as the baseline specification, while the less restrictive process more than doubles the number of observations.

Since there is no single "treatment year" for all observations, simple matching was done by pairing block groups that were treated at any time between 2015 and 2020 with those block groups that never met the 50 % coverage required for each threshold for any of the years in the sample. It should be noted that CEM may significantly reduce sample size, since block groups that do not match with the chosen covariates are left out of the analysis. For example, 100 Mbps service was relatively common during the study period, and block groups lacking this level of availability (potential controls) likely had "lower" bin signatures (i.e., lower bins in race or education). Treated block groups with "higher" bin signatures are therefore less likely to be matched with controls – and thus are dropped from the analysis. Our panel dataset comprises eight states and more than 12,000 block groups (after dropping those with 100 Mbps by 2014), so this reduction of sample size was not a serious concern when considering the full dataset.

Using the CEM matched samples, we compare census block groups that had access to specific internet speeds from 2015 to 2020 with those that only had access to lesser speeds. Specifically, we observe the effect that three levels of "fast" internet speeds have on business births and look at whether and how the effects change based on the speed considered. We estimate the event study laid out in Eq. (2) using the user-written Stata package *eventdd* (Clarke and Schythe, 2020). The *eventdd* package estimates and plots the estimated effects, providing a clear visualization of effects over time. After considering the overall effect, we look at subsets of the data to assess the heterogeneity of effects between metro and non-metro block groups and by industry.

To address the issue of high-frequency zeros in the dependent variable in some cases, we also run panel Poisson models using the underlying count data rather than births per 10,000 population as in the original specification. Since the data is not censored, a Tobit model is not appropriate. We choose a Poisson with robust standard errors over a negative binomial model because the data are not dramatically overdispersed. In fact, for the industry subsets, the mean of the dependent variable is typically slightly larger than the variance (i.e., underdispersed). Under these circumstances, Poisson is the more appropriate choice and is beneficial in that it requires no assumptions about the conditional variance (Nichols, 2010; Wooldridge, 2020).

Table 5
CEM matching covariates by model specification.

Variable	Base	More Restrictive	Less Restrictive
Population (2014–2018)	X	X	X
% non-white	X	X	
% bachelor's degree or higher	X	X	
% households with broadband connection	X		
Total businesses	X		X
Goods-producing share of businesses	X		
Baseline (2015) download speed (Mbps)		X	
% in poverty		X	X
Median Household Income		X	X
% employed		X	X
Rural-urban Continuum Code (1–9)			X
# Block Groups (100 MBPS vs. < 100)	3,437	2,204	8,118

Results

The CEM results indicate that matching was successful at decreasing the imbalance for almost all variables in each model specification.[22] Table 6 displays the measures of imbalance between the treated and control block groups before and after the baseline CEM, as well as the

Table 6
Measures of covariate imbalance in pre- and post-matching, full sample.

L_1 Measures	> 50 % 100+ Mbps		> 50 % 250 Mbps		> 50 % 1000 Mbps	
	Initial	Post-CEM	Initial	Post-CEM	Initial	Post-CEM
Population 2014–18	0.133	0.098	0.155	0.083	0.068	0.059
% non-white	0.266	0.000	0.453	0.020	0.199	0.018
% bachelor's degree or higher	0.202	0.078	0.195	0.136	0.037	0.061
% households with broadband connection	0.347	0.106	0.270	0.106	0.071	0.047
Total businesses in block group	0.258	0.167	0.179	0.183	0.037	0.085
Goods-producing ratio	0.257	0.043	0.289	0.042	0.076	0.039
# Obs – Treated	11,670	2,276	3,529	522	6,199	2,079
# Obs – Control	1,997	1,161	1,942	420	3,529	1,662
Total # Obs	13,667	3,437	5,471	942	9,728	3,741

Note: L_1 measures the histogram difference between treated and control group for a particular variable. L_1 takes on a value between 0 and 1 with higher values indicating a larger imbalance.

[20] For more details on CEM see Iacus et al. (2009, 2011, 2012). Studies across the natural and social sciences have noted the benefits of CEM over other matching methods or have compared CEM to other matching methods (e.x., King and Nielsen, 2019; Bertoni et al., 2020; and Ripollone et al., 2020).

[21] We use the default binning algorithm within Stata's *cem* command which generates 17 "bins" for each of the covariates. The default binning method within the *cem* package is Sturge's Rule which uses the following formula to determine the optimal number of bins: *Optimal bins* $= \lceil \log_2 n + 1 \rceil$, where n is the total number of observations in the dataset.

[22] A general reduction from the original imbalance may be more important than reaching a specific threshold (Iacus et al. 2011; 2012). However, it has also been argued that 0.10 is an appropriate threshold to aim for (Austin, 2009).

resulting reduction in sample size for the base specification. The matched samples consist of 17.2 % to 38.5 % of the total block groups in the initial data. CEM results using the more and less restrictive specifications similarly improve imbalance across nearly all matching covariates.

Fig. 3 displays the graphical *eventdd* output for the main analysis. Each chart represents results for one of the three speed treatment thresholds for the matched sample. Each coefficient estimate is indicated by a solid dot, with 95 % confidence interval bars displayed. The vertical black line at Time = −1 represents the first lead (i.e. immediately prior to treatment) which is treated as the default (omitted) category.

The coefficients for every lead period (to the left of the vertical black lines) for the first two speed thresholds in Fig. 3 (100+ Mbps; 250 Mbps) are not statistically different from zero. This indicates that in the years leading up to a block group being treated, births per 10,000 population were not significantly different between treatment and control groups in the matched sample. This suggests that the parallel trends assumption, the primary identifying assumption of this model, holds. Conversely, the parallel trends assumption does not appear to hold for gigabit service, where treated block groups had lower levels of births prior to the introduction of this very high-speed broadband.

In the periods *after* a block group is treated, births per 10,000 population exhibit an upward sloping trend and are statistically significant for block groups treated at the 100+ Mbps and 250 Mbps levels. Recall that the relevant control groups for these regressions were block groups with less than 100 Mbps and those with 100 Mbps (but not 250), respectively. The gigabit versus 250 Mbps event-study plot similarly suggests an increase in births after intervention; however, the pre-treatment trends are problematic and detract from any potential claims of causality. In fact, robustness checks changing the definition of treatment from 50 % of establishments in Fig. 3 to 75 % or 90 % show even stronger results for the introduction of 100+ Mbps and 250 Mbps, but no significant impacts for the introduction of gigabit service (Appendix B).

These results provide causal evidence that fast – but perhaps not “ultra-fast” speeds – were important to business start-ups during the period studied here. This finding is in line with previous results suggesting that a speed threshold may exist, beyond which further quality improvements may be unproductive (Koutroumpis, 2019; Deller et al., 2022).

Table 7 displays the numeric regression results that correspond to Fig. 3. Given the scaling of the dependent variable (births per 10,000 population), it is helpful to consider what these estimated coefficients might imply for a tangible geography such as a small metro area with a population of 50,000 or a major metropolitan area with a population of one million or more. For example, the cumulative effect for years 1–5 for the 250 Mbps treated block groups is 0.0614 births per 10,000 population. In a small metro area with 50,000 people, this amounts to 0.31 business births and in a major metropolitan area with 1 million people, this amounts to 6.14 business births. The population of 52.2 million for the eight states studied here suggests an increase of roughly 320 business births *above and beyond* what might be expected with only 100 Mbps available. Results for the more and less restrictive matching specifications can be found in Appendix C and generally support those in Table 7. In particular, the pre-trends for the gigabit treated group remain problematic.

Next, we consider the heterogeneity of effects in metro versus non-metro block groups. Matching (via CEM) was re-done within the metro and non-metro subsamples, respectively. Improvements in the covariate imbalances and the final matched sample sizes can be found in Appendix D. Fig. 4 displays the graphical *eventdd* output for each of the three broadband treatment speeds. The top row of the figure displays results for the metro subsample while the bottom row contains the non-

Table 7
Regression of business births per 10,000 population on timing of highspeed broadband access, full sample.

	100+ Mbps (vs < 100)	250 Mbps (vs 100)	Gigabit (vs 250)
lead4	−0.0000671	−0.000371	−0.0183***
	(0.0009)	(0.0034)	(0.0036)
lead3	0.000553	−0.00347*	−0.0162***
	(0.0007)	(0.0021)	(0.0033)
lead2	0.00044	−0.0012	−0.0038
	(0.0006)	(0.0015)	(0.0029)
lead1	Reference		
lag0	0.00303***	0.0024	0.00888***
	(0.0012)	(0.0018)	(0.0027)
lag1	0.00500***	0.00932***	0.0134***
	(0.0007)	(0.0023)	(0.0028)
lag2	0.00736***	0.00949***	0.0254***
	(0.0011)	(0.0036)	(0.0060)
lag3	0.0125***	0.00637	0.0314***
	(0.0015)	(0.0040)	(0.0054)
lag4	0.0125***	0.0188**	0.0350***
	(0.0020)	(0.0088)	(0.0077)
lag5	0.0149***	0.0238**	0.0361***
	(0.0037)	(0.0108)	(0.0059)
Constant	0.00554***	0.0114***	0.0372***
	(0.0005)	(0.0007)	(0.0019)
Observations	20,622	5,652	22,446
R-squared	0.698	0.615	0.648

Notes: This table reports event study estimates of the effect of broadband internet speed on business births per 10,000 population at the census block group level for eight U.S. states. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed. Corresponding control groups are made up of block groups which have service at the next lower speed threshold only (i.e., 100+ Mbps vs < 100 Mbps, 250 Mbps vs 100 Mbps, Gigabit vs 250 Mbps). The lead estimates indicate years prior to treatment. The year prior to receiving treatment is omitted (Lead1). Lag estimates indicate years since treatment. Robust standard errors clustered at the block group level are in parentheses. Asterisks ***, **, and * indicate statistical significance at the 1 %, 5 %, and 10 % levels.

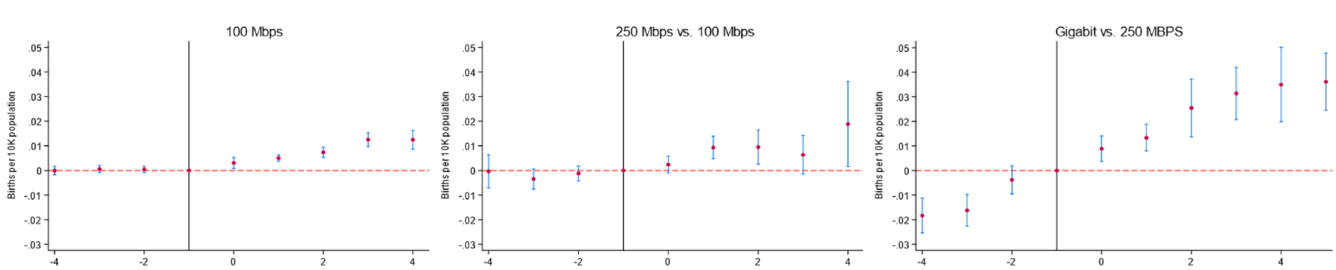

Fig. 3. Regression of business births per 10,000 population on timing of highspeed broadband access, full sample.
Notes: Panels display the event study estimation for the Coarsened Exact Matching (CEM) matched sample. Point estimates are displayed with their 95 % confidence intervals as described in Eq. (2). The baseline (omitted) period is one year prior to treatment in each block group, indicated by the solid vertical line in each plot. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed.

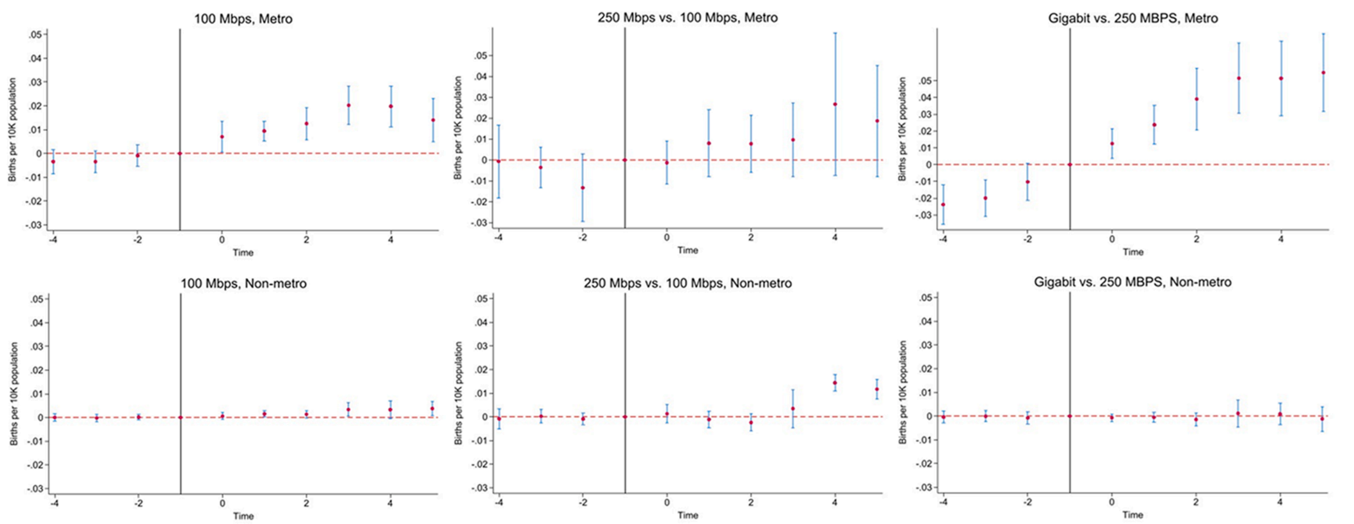

Fig. 4. Regression of business births per 10,000 population on timing of highspeed broadband access, Metro (top) and non-Metro (bottom).
Notes: Panels display the event study estimation for metro and non-metro block groups, respectively. Point estimates are displayed with their 95 % confidence intervals as described in Eq. (2). The lead estimates indicate years prior to treatment. The baseline (omitted) period is one year prior to treatment, indicated by the solid vertical line in each plot. Lag estimates indicate years since treatment. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed.

metro results. Similar to the aggregate results, coefficients on the lead time periods are almost all statistically equal to zero which supports the parallel trends assumption. The only exception to this is the Gigabit vs. 250 Mbps sample for metro locations. For metro block groups treated with 100 Mbps+, a statistically significant effect is seen for all lag periods with the greatest effect of around 0.02 births per 10,000 in lags 3 and 4. However, no significant effects are observed for the Metro block groups treated with 250 Mbps. Metro block groups treated with gigabit speed see the largest effect of around 0.05 births per 10,000 in lags 3, 4, and 5; but the pre-treatment trends remain problematic.

In non-metro areas, the impacts of obtaining better broadband speeds are smaller but still present. Non-metro block groups treated at the 100+ Mbps level observe statistically significant increases in business births in all lag years: around 0.001 births per 10,000 population in lags 1 and 2, and around 0.003 in lags 3, 4, and 5 – an order of magnitude below those in metropolitan locations. Interestingly, the aggregate results for 250 Mbps (Fig. 3) appear to be driven by non-metro locations, with strongly significant results in lags 4 and 5. No impact is observed for gigabit speeds. These coefficients imply that improved broadband speeds are not likely to lead to dramatic increases in business creation for rural geographies; however, the overstatement of the broadband availability data is more problematic in rural locations (Ford, 2022) and may play a role in these results. Numeric regression results for these specifications can be found in Appendix E.

As prior research has pointed out, the effects of broadband may vary by industry. Appendix F contains event study estimates (both plots and tables) of the effect of 100+ Mbps, 250 Mbps and gigabit broadband internet speed on births per 10,000 population in the eight industries covered by the data. In line with the overall results, the introduction of 100+ Mbps has strongly statistically significant effects on business births in each of the eight industries considered here, though the level of significance varies by industry and by time (lag). Additionally, there are very few significant lead periods, suggesting that the parallel trends assumption holds. For block groups gaining 100+ Mbps service (relative to low- to no broadband at all), these results support the argument that this new access facilitates business activity. The introduction of 250 Mbps (vs. 100 Mbps) also shows robust but less universal results, with significant lags at the $p < 0.05$ level in five of the eight industries and again minimal significance in the periods prior. Interestingly, some of the 100+ Mbps results show up in the year service was added (retail, real estate, professional services, and food / accommodations) while most of the 250 Mbps results first appear one year after. Gigabit service, compared against 250 Mbps service, again shows strong positive impacts on births with statistically significant lags in all eight industries. However, all eight also demonstrate at least some pre-trends that are problematic for claims of causality.

Additional robustness checks

The analysis to this point has already demonstrated that the baseline results generally hold when the threshold for "treatment" moves from 50 % of businesses with broadband access to 75 % or 90 % (Appendix B). It has also demonstrated that the results largely hold when considering more or less restrictive matching covariates (Appendix C). We turn now to three additional robustness checks: the potential influence of COVID-19; the use of more recent estimators addressing treatment timing variation; and Poisson modeling for industry subsets due to high frequency of zeros.

The 2015–2020 period of analysis overlaps with the COVID-19 pandemic, and it is possible that the inclusion of 2020 is skewing the results. For example, it is well documented that applications for business births surged in late 2020 (Decker and Haltiwanger, 2022). We address this by re-running the baseline regressions but omitting the year 2020. The results (Appendix G) show that the takeaway findings remain, even when only up to four lag years can be included (i.e. when lag0 = 2015).

The TWFE approach used in our baseline analysis may produce misleading results due to treatment heterogeneity between groups or over time, particularly in settings such as this where treatment timing varies. Alternative empirical approaches have been proposed that deal with this problem in different ways (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; de Chaisemartin and D'Haultfoeuille, 2022). Clarke and Schythe (2020) suggest verifying the robustness of the event study methodology used here by implementing one or more of these approaches. The results, using the Callaway and Sant'Anna methodology and the *csdid* command in Stata, verify that our baseline results hold for 100+ Mbps and 250 Mbps (Appendix H). This includes the positive upward trend post-event for 100+ Mbps and statistical significance in only years 1 and 2 for 250 Mbps, both of which are very similar to Fig. 3. The gigabit event-study plot generally lacks significance in the lag periods, suggesting that the treatment timing issue may be more problematic for this subset and reinforcing the lack of robust empirical evidence supporting gigabit speeds' influence on business births when compared to a speed threshold of 250 Mbps. The agreement of this more robust methodology gives more credence to claims of causality.

Table 2 demonstrated that the industry-specific births contain a high percentage of zero entries and may not be appropriate for OLS regression. We run Poisson fixed effects regression with robust standard errors using the count of businesses per block group (not per 10,000 population) and report these industry-specific regression results in Appendix I.[23] These results are also generally supportive of the event-study findings. There are statistically significant lag results in five of the eight industries for 100+ Mbps (compared to eight out of eight in Appendix D). The results for 250 Mbps are even more closely aligned, with both methods agreeing on the importance of this higher-speed threshold for manufacturing, retail trade, finance, and food / accommodations. Interestingly, information sector births appear to be influenced by the addition of 100+ Mbps, while the finance sector is more heavily impacted by 250 Mbps. This suggests that the different technologies or data needs in these industries require distinct bandwidth thresholds. The gigabit results continue to find problematic pre-trends in industries where post-event significance is found.

Conclusion

Our general findings show that 100+ Mbps availability leads to increases in business births per capita across our 8-state sample, and that there is a benefit associated with 250 Mbps service over 100 Mbps. There is noticeable heterogeneity between rural and urban areas, however, with treatment effects that are much more pronounced in metropolitan locations. This is the case *specifically for the 2015–2020 period* and does not necessarily reflect forthcoming technological developments. As Koutroumpis (2019) also argues, this threshold is likely to increase over time as more applications and skills require additional speed. Emerging technologies including artificial intelligence, virtual reality, and the internet of things (IoT) require considerably faster broadband speed and more bandwidth (Goepel, 2018; Ayar Labs, 2022). This is also true for rural-focused fields like precision agriculture, where machine / deep learning and real-time data processing rely on higher-level broadband networks (Karunathilake et al., 2023; Shaikh, Rasool, and Lone, 2022). A recent study by Han et al. (2023) suggests that cloud computing is another emerging technology whose use is impacted by the availability of broadband. As these technologies continue to evolve and make their way into normal business processes, we would expect to see business births respond to the need for additional speed.

There are several limitations to this analysis. First, this study is restricted to only eight U.S. states. Future extensions of this work may consider additional states or look more closely at effects within individual states, rather than a region in aggregate as done here. Another limitation is that the establishment-level data only covered eight out of 20 two-digit NAICS industries. Though the eight industries considered here were chosen specifically for their likely use of or reliance on internet technologies, it is possible that broadband may affect industries not studied here. In particular, recent advances in artificial intelligence could have impacts on the educational services or arts and entertainment sectors that might only be realized with high-speed connections. Third, the business establishment data used in this study was collected by a private company which, to our knowledge, has not yet been rigorously compared to other public sector or private datasets such as the Bureau of Labor Statistic's Quarterly Census of Employment and Wages (QCEW) or the National Establishment Time Series (NETS).

Recent U.S. legislation has prioritized building highspeed broadband infrastructure, and over 11 million physical locations across the U.S. have been identified as lacking broadband at 100/20 Mbps speeds (Conlow, 2023). Some internet service providers have already gone to great lengths to install fiber optic cable with speeds of one gigabit or more. This study sought to identify whether a causal relationship exists between highspeed broadband internet availability and business births. The treatment threshold – greater than 50 % of establishments within a block group having access to a specific broadband speed – is reached at different points in time for each block group, making an event study design appropriate. Our results suggest that the current legislation's emphasis on increasing at least 100 Mbps availability is justified, particularly in metro block groups and in the eight specific industries considered here. Speeds of 100+ Mbps also had a positive and significant (but smaller) effect on business births in non-metro areas. Since the IIJA is focused on bringing at least 100 Mbps service to all Americans, and rural locations in particular, these results suggest that there could be immediate impacts in terms of businesses added in the near future.

Ultimately, we find no strong evidence that offering gigabit service resulted in additional business creation during our period of analysis. We do find evidence that pushing beyond 100 Mbps access to 250 Mbps contributed to incremental business births, particularly in rural areas, but require longer lag times to accrue. As business technologies requiring faster high-speed internet access advance and make their way into mainstream business operations, there may well be additional shifts in business creation dependent on broadband internet at these higher speeds.

Author statement

Author statement: All procedures in this manuscript were performed in compliance with relevant laws and institutional guidelines. All data used in this manuscript was secondary in nature and did not require approval from an institutional review board (IRB).

Funding

This work was partially supported by the intramural research program of the U.S. Department of Agriculture, National Institute of Food and Agriculture, Hatch Project 1018233 and the Neustadt Chair at Oklahoma State University.

CRediT authorship contribution statement

Christina Biedny: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Brian E. Whitacre:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Validation. **Andrew J. Van Leuven:** Conceptualization, Data curation, Validation, Writing – original draft, Writing – review & editing.

Data availability

The data that has been used is proprietary.

APPENDICES

[23] Note that the number of observations is lower in the Poisson regressions than those in Appendix F because block groups with all zero outcomes are omitted.

Appendix A. Stylized example of lead and lag variables used in the event study estimation

Block Group (b)	Year (t)	Year Treated	Post Event	Relative Year	Lead 4	Lead 3	…	Lag 0	Lag 1	…	Lag 5
A	2015	2016	0	−1	0	0	•••	0	0	•••	0
A	2016	2016	1	0	0	0	•••	1	0	•••	0
A	2017	2016	1	1	0	0	•••	0	1	•••	0
A	2018	2016	1	2	0	0	•••	0	0	•••	0
A	2019	2016	1	3	0	0	•••	0	0	•••	0
A	2020	2016	1	4	0	0	•••	0	0	•••	0
B	2015	2019	0	−4	0	1	•••	0	0	•••	0
B	2016	2019	0	−3	0	0	•••	0	0	•••	0
B	2017	2019	0	−2	0	0	•••	0	0	•••	0
B	2018	2019	0	−1	0	0	•••	0	0	•••	0
B	2019	2019	1	0	0	0	•••	1	0	•••	0
B	2020	2019	1	1	0	0	•••	0	1	•••	0
C	2015	•	0	•	0	0	•••	0	0	•••	0
C	2016	•	0	•	0	0	•••	0	0	•••	0
C	2017	•	0	•	0	0	•••	0	0	•••	0
C	2018	•	0	•	0	0	•••	0	0	•••	0
C	2019	•	0	•	0	0	•••	0	0	•••	0
C	2020	•	0	•	0	0	•••	0	0	•••	0
D	2015	2015	1	0	0	0	•••	1	0	•••	0
D	2016	2015	1	1	0	0	•••	0	1	•••	0
D	2017	2015	1	2	0	0	•••	0	0	•••	0
D	2018	2015	1	3	0	0	•••	0	0	•••	0
D	2019	2015	1	4	0	0	•••	0	0	•••	0
D	2020	2015	1	5	0	0	•••	0	0	•••	1

Notice that for untreated block groups (block group C), *Year Treated* is omitted and as such, *Relative Year* is missing for these true control units. This results in all of the binary *Lead* and *Lag* variables taking the value zero. The *eventdd* package in Stata uses *Relative Year* to automatically generate the binary *Lead* and *Lag* dummies used in estimation. Broadband availability data for the year 2014 was used to inform whether a block group was identified as treated in 2015 (block group D). Block groups with 100 Mbps service or better by 2014 were omitted from the analysis.

Appendix B. Robustness Check using 75 % and 90 % thresholds for Treatment

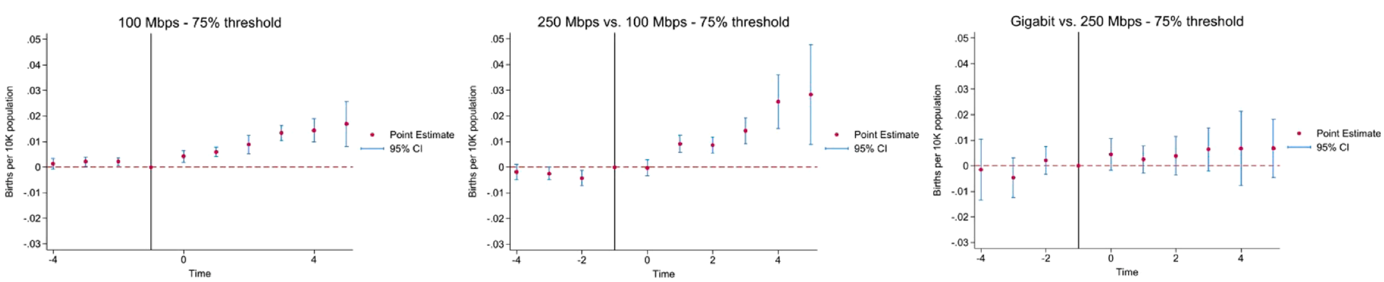

Fig. B.1. Regression of business births per 10,000 population on timing of highspeed broadband access, full sample, 75 % Treatment threshold.
Notes: Panels display the event study estimation for the baseline Coarsened Exact Matching (CEM) matched sample. Point estimates are displayed with their 95 % confidence intervals as described in Eq. (2). The baseline (omitted) period is one year prior to treatment in each block group, indicated by the solid vertical line in each plot. ***Treatment is defined as greater than or equal to 75 % of establishments in a block group having access*** to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed.

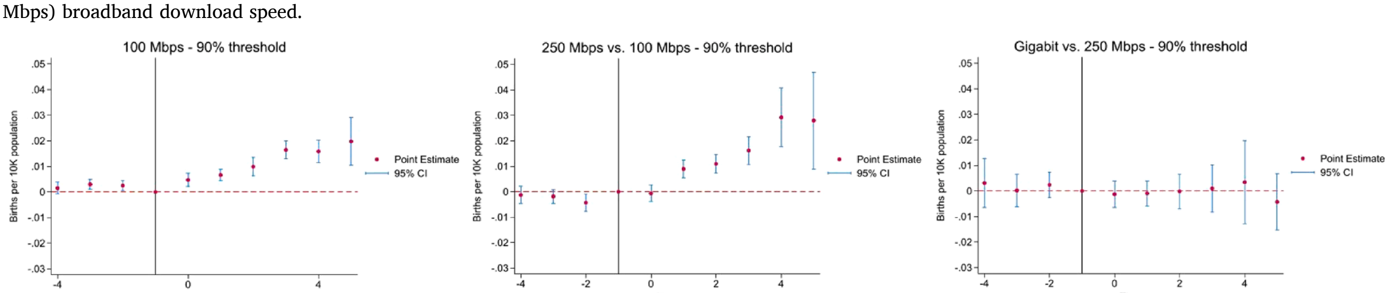

Fig. B.2. Regression of business births per 10,000 population on timing of highspeed broadband access, full sample, 90 % Treatment threshold.
Notes: Panels display the event study estimation for the baseline Coarsened Exact Matching (CEM) matched sample. Point estimates are displayed with their 95 % confidence intervals as described in Eq. (2). The baseline (omitted) period is one year prior to treatment in each block group, indicated by the solid vertical line in each plot. ***Treatment is defined as greater than or equal to 90 % of establishments in a block group having access*** to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed.

Appendix C. Robustness Check using more and less restrictive CEM matching variables

Table C.1
Regression of business births per 10,000 population on timing of highspeed broadband access, full sample – More Restrictive Specification.

	100+ Mbps (vs < 100)	250 Mbps (vs 100)	Gigabit (vs 250)
lead4	−0.0008	0.0002	−0.0115***
	(0.0008)	(0.0023)	(0.0022)
lead3	0.0007	−0.0020	−0.0107***
	(0.0006)	(0.0021)	(0.0039)
lead2	0.0005	−0.0030	−0.0040
	(0.0006)	(0.0025)	(0.0031)
lead1	Reference		
lag0	0.0036***	0.0040	0.0092*
	(0.0008)	(0.0029)	(0.0047)
lag1	0.0042***	0.0090***	0.0080**
	(0.0008)	(0.0033)	(0.0033)
lag2	0.0049***	0.0066**	0.0156***
	(0.0008)	(0.0031)	(0.0040)
lag3	0.0109***	−0.0016	0.0318***
	(0.0014)	(0.0042)	(0.0068)
lag4	0.0088***	0.0031	0.0160***
	(0.0015)	(0.0035)	(0.0044)
lag5	0.0114***	omitted	0.0243***
	(0.0033)		(0.0068)
Constant	0.0045***	0.0096***	0.0234***
	(0.0003)	(0.0008)	(0.0013)
Observations	13,224	2,772	6,936
R-squared	0.6023	0.6674	0.6295

Notes: This table reports event study estimates of the effect of broadband internet speed on business births per 10,000 population at the census block group level for eight U.S. states. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed. Corresponding control groups are made up of block groups which have service at the next lower speed threshold only (i.e., 100+ Mbps vs < 100 Mbps, 250 Mbps vs 100 Mbps, Gigabit vs 250 Mbps). Covariates used to conduct matching include population,% non-white,% with bachelor's degree or higher, baseline (2015) download speed,% in poverty, Median Household Income, and% employed. The lead estimates indicate years prior to treatment. The year prior to receiving treatment is omitted (Lead1). Lag estimates indicate years since treatment. Robust standard errors clustered at the block group level are in parentheses. Asterisks ***, **, and * indicate statistical significance at the 1 %, 5 %, and 10 % levels.

Table C.2
Regression of business births per 10,000 population on timing of highspeed broadband access, full sample – Less Restrictive Specification.

	100+ Mbps (vs < 100)	250 Mbps (vs 100)	Gigabit (vs 250)
lead4	0.00292**	−0.00683***	−0.0200***
	(0.0012)	(0.0020)	(0.0031)
lead3	0.00455***	−0.00653***	−0.0149***
	(0.0009)	(0.0013)	(0.0027)
lead2	0.00431***	−0.00659***	−0.00586**
	(0.0009)	(0.0015)	(0.0024)
lead1	Reference		
lag0	0.00822***	−0.00355**	0.0112***
	(0.0013)	(0.0016)	(0.0026)
lag1	0.0130***	0.0113***	0.0198***
	(0.0012)	(0.0020)	(0.0026)
lag2	0.0160***	0.0130***	0.0292***
	(0.0021)	(0.0019)	(0.0042)
lag3	0.0251***	0.0162***	0.0441***
	(0.0021)	(0.0029)	(0.0050)
lag4	0.0259***	0.0370***	0.0444***
	(0.0024)	(0.0087)	(0.0051)
lag5	0.0268***	0.0330***	0.0455***
	(0.0025)	(0.0069)	(0.0053)
Constant	0.00860***	0.0176***	0.0415***
	(0.0010)	(0.0006)	(0.0017)
Observations	48,708	16,926	43,446
R-squared	0.656	0.643	0.684

Notes: This table reports event study estimates of the effect of broadband internet speed on business births per 10,000 population at the census block group level for eight U.S. states. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed. Corresponding control groups are made up of block groups which have service at the next lower speed threshold only (i.e., 100+ Mbps vs < 100 Mbps, 250 Mbps vs 100 Mbps, Gigabit vs 250 Mbps). Covariates used to conduct matching include population, total number of businesses,% in poverty, Median Household Income,% employed, and Rural-Urban Continuum Code (RUCC). The lead estimates indicate years prior to treatment. The year prior to receiving treatment is omitted (Lead1). Lag estimates indicate years since treatment. Robust standard errors clustered at the block group level are in parentheses. Asterisks ***, **, and * indicate statistical significance at the 1 %, 5 %, and 10 % levels.

Appendix D. Coarsened Exact Matching (CEM) impacts on covariate imbalance and sample size, Metro and non-Metro subsamples

Table D.1
Measures of covariate imbalance in pre- and post-matching, Metro.

L_1 Measures	> 50 % 100+ Mbps		> 50 % 250 Mbps		> 50 % 1000 Mbps	
	Initial	Post-CEM	Initial	Post-CEM	Initial	Post-CEM
Population 2014–18	0.124	0.119	0.122	0.106	0.071	0.071
% non-white	0.320	0.037	0.526	0.020	0.179	0.000
% bachelor's degree or higher	0.260	0.131	0.191	0.165	0.057	0.067
% households with broadband connection	0.412	0.085	0.280	0.150	0.083	0.095
Total bbusinesses in block group	0.296	0.187	0.210	0.188	0.045	0.071
Goods-producing ratio	0.342	0.049	0.330	0.003	0.059	0.058
# Obs – Treated	7,741	431	2,774	254	4,108	1,526
# Obs – Control	714	236	859	136	2,774	1,313

Notes: L_1 measures the histogram difference between treated and control group for a particular variable.
L_1 takes on a value between 0 and 1 with higher values indicating a larger imbalance.

Table D.2
Measures of covariate imbalance in pre- and post-matching, non-Metro.

L_1 Measures	> 50 % 100+ Mbps		> 50 % 250 Mbps		> 50 % 1000 Mbps	
	Initial	Post-CEM	Initial	Post-CEM	Initial	Post-CEM
Population 2014–18	0.070	0.069	0.100	0.097	0.052	0.086
% non-white	0.126	0.000	0.210	0.031	0.151	0.029
% bachelor's degree or higher	0.069	0.097	0.144	0.109	0.077	0.035
% households with broadband connection	0.233	0.095	0.175	0.123	0.082	0.047
Total businesses in block group	0.247	0.160	0.167	0.094	0.053	0.087
Goods-producing ratio	0.132	0.080	0.170	0.063	0.106	0.040
# Obs – Treated	3,929	568	755	32	2,091	260
# Obs – Control	1,283	431	1,083	38	755	203

Notes: L_1 measures the histogram difference between treated and control group for a particular variable.
L_1 takes on a value between 0 and 1 with higher values indicating a larger imbalance.

Appendix E. Metro and non-Metro subsample regression analysis

Table E.1
Regression of business births per 10,000 population on timing of highspeed broadband access, Metro and non-Metro subsamples.

	100+ Mbps (vs. <100)		250 Mbps (vs. 100)		Gigabit (vs. 250)	
	Metro	non-Metro	Metro	non-Metro	Metro	non-Metro
lead4	−0.0034	0.00000192	−0.000697	−0.0008	−0.0237***	−0.000364
	(0.0026)	(0.0008)	(0.0089)	(0.0021)	(0.0060)	(0.0013)
lead3	−0.00341	−0.000201	−0.00358	0.000265	−0.0199***	−0.0000323
	(0.0023)	(0.0007)	(0.0049)	(0.0014)	(0.0055)	(0.0012)
lead2	−0.000864	0.000154	−0.0133	−0.000867	−0.0103*	−0.000773
	(0.0023)	(0.0006)	(0.0083)	(0.0013)	(0.0056)	(0.0013)
lead1	Reference					
lag0	0.00704**	0.000593	−0.0013	0.00132	0.0126***	−0.000679
	(0.0033)	(0.0007)	(0.0052)	(0.0019)	(0.0046)	(0.0008)
lag1	0.00946***	0.00153**	0.008	−0.00111	0.0238***	−0.00052
	(0.0021)	(0.0007)	(0.0081)	(0.0017)	(0.0058)	(0.0011)
lag2	0.0126***	0.00137*	0.00775	−0.00239	0.0391***	−0.00142
	(0.0034)	(0.0007)	(0.0070)	(0.0018)	(0.0094)	(0.0014)
lag3	0.0203***	0.00336**	0.00966	0.00349	0.0515***	0.00118
	(0.0041)	(0.0014)	(0.0090)	(0.0040)	(0.0106)	(0.0029)
lag4	0.0198***	0.00331*	0.0267	0.0144***	0.0514***	0.000944
	(0.0044)	(0.0019)	(0.0173)	(0.0018)	(0.0113)	(0.0023)
lag5	0.0141***	0.00380***	0.0187	0.0117***	0.0548***	−0.00116
	(0.0046)	(0.0015)	(0.0135)	(0.0021)	(0.0118)	(0.0026)
Constant	0.00887***	0.00312***	0.0224***	0.00283***	0.0566***	0.00651***
	(0.0014)	(0.0004)	(0.0020)	(0.0005)	(0.0030)	(0.0005)
Observations	4,002	5,994	2,340	420	17,034	2,778
R-squared	0.529	0.553	0.68	0.382	0.676	0.603

Notes: This table reports event study estimates of the effect of broadband internet speed on business births per 10,000 population at the Census block group level for eight U.S. states. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed. The lead estimates indicate years prior to treatment. The year prior to receiving treatment is omitted (Lead1). Lag estimates indicate years since treatment. Robust standard errors clustered at the block group level are in parentheses. Asterisks ***, **, and * indicate statistical significance at the 1 %, 5 %, and 10 % levels.

Appendix F. Regression of births per 10,000 population on timing of highspeed broadband access, by industry

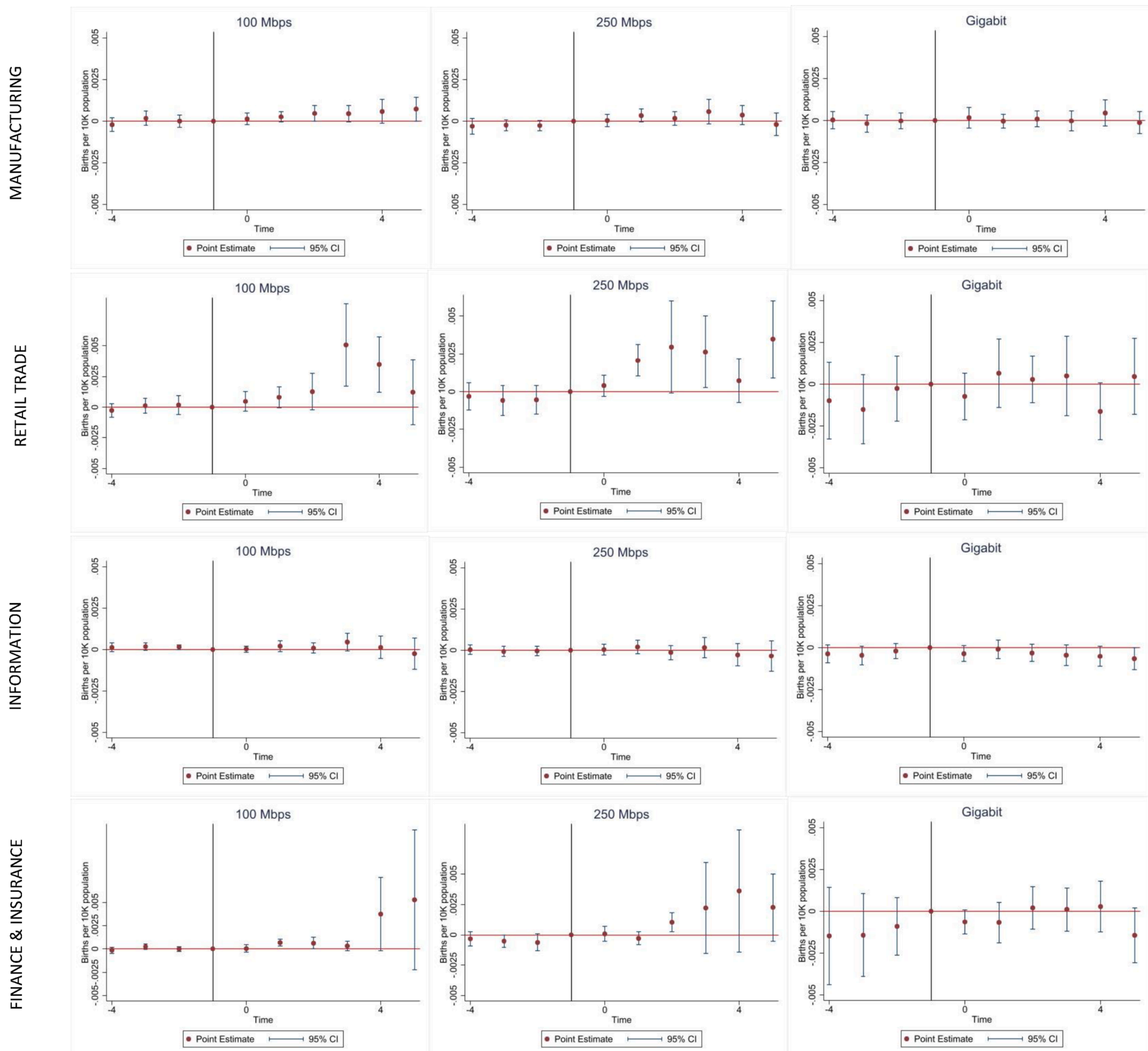

Fig. F.1. Regression of births per 10,000 population on timing of highspeed broadband access, by industry.
Notes: Panels display the event study estimation for block groups by industry. Point estimates are displayed with their 95 % confidence intervals as described in Eq. (2). The lead estimates indicate years prior to treatment. The baseline (omitted) period is one year prior to treatment, indicated by the solid vertical line in each plot. Lag estimates indicate years since treatment. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed.

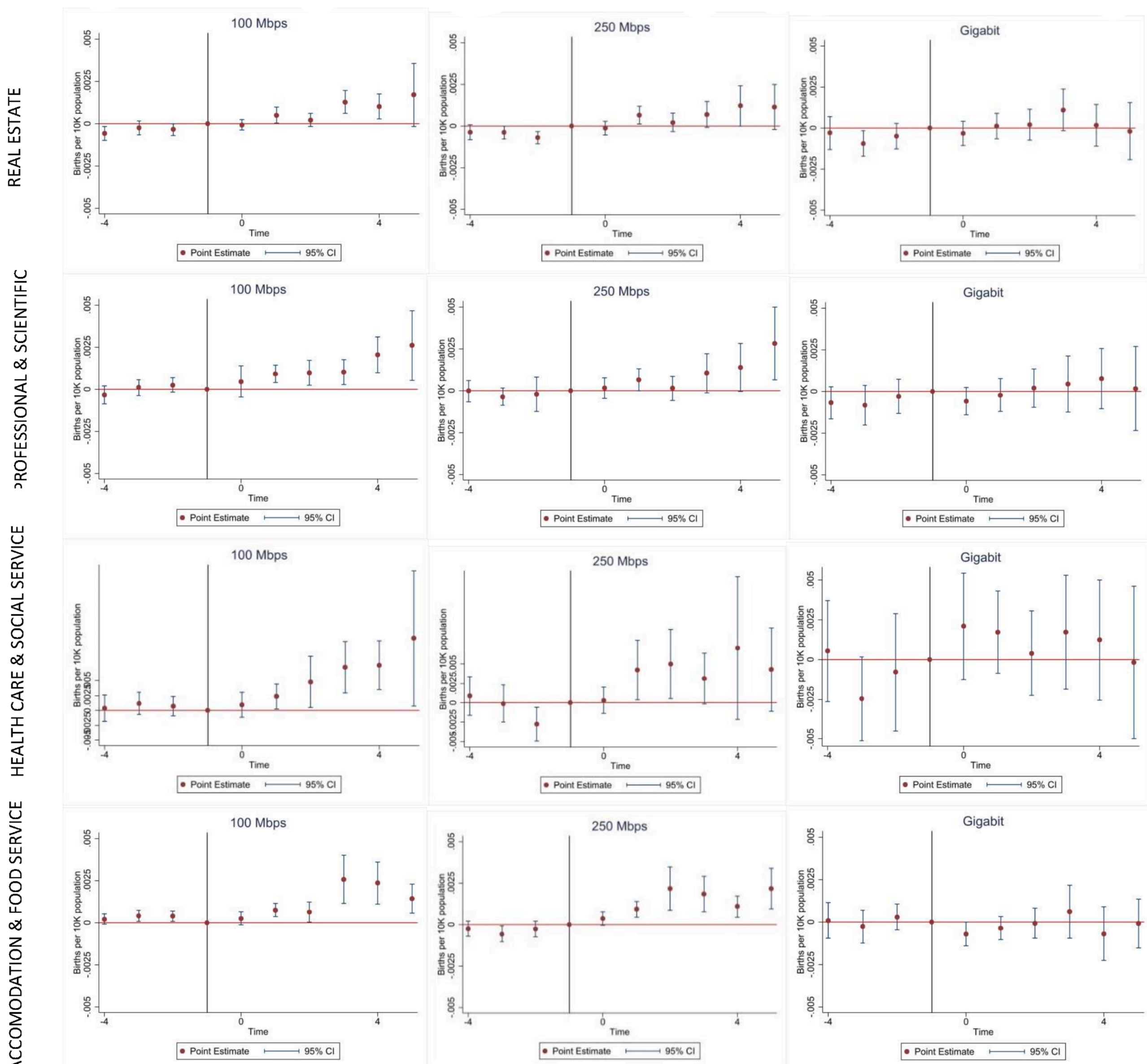

Fig. F.1. (*continued*).

Table F.1
Regression of births per 10,000 population on timing of highspeed broadband access, by industry.

	100+ Mbps (vs. <100)				250 Mbps (vs. 100)				Gigabit (vs. 250)			
	Manufacturing	Retail Trade	Information	Finance	Manufacturing	Retail Trade	Information	Finance	Manufacturing	Retail Trade	Information	Finance
lead4	0.00001	−0.00002	0.00000	−0.000310**	−0.00003	0.00052	0.00024	0.00016	−0.00005	−0.00308***	−0.000737***	−0.00160***
	(0.00010)	(0.00030)	(0.00006)	(0.00016)	(0.00027)	(0.00066)	(0.00024)	(0.00059)	(0.00026)	(0.00054)	(0.00023)	(0.00048)
lead3	0.00014	0.000481*	−0.00001	−0.00010	0.00001	0.00012	0.00020	−0.000577**	−0.00017	−0.00259***	−0.00024	−0.00050
	(0.00012)	(0.00025)	(0.00005)	(0.00016)	(0.00020)	(0.00054)	(0.00017)	(0.00025)	(0.00022)	(0.00053)	(0.00017)	(0.00055)
lead2	0.00001	0.00023	0.00004	−0.00005	0.00003	−0.00057	0.00016	0.00009	−0.00001	−0.00051	−0.00030	−0.00032
	(0.00009)	(0.00023)	(0.00006)	(0.00012)	(0.00020)	(0.00051)	(0.00012)	(0.00037)	(0.00020)	(0.00063)	(0.00026)	(0.00065)
lead1	Reference											
lag0	0.000202*	0.000817**	0.00006	0.00018	0.00060	0.00074	0.00019	0.000716*	0.000530*	0.00120**	0.000557*	0.00138**
	(0.00011)	(0.00035)	(0.00006)	(0.00015)	(0.00036)	(0.00059)	(0.00018)	(0.00038)	(0.00032)	(0.00054)	(0.00034)	(0.00061)
lag1	0.00004	0.000871***	0.00012	0.000540***	0.00111***	0.00260***	0.00113**	0.00033	0.000567**	0.00214***	0.000704**	0.00181**
	(0.00010)	(0.00028)	(0.00008)	(0.00015)	(0.00041)	(0.00095)	(0.00058)	(0.00039)	(0.00025)	(0.00059)	(0.00032)	(0.00092)
lag2	0.000318*	0.00136***	0.000209*	0.000765***	0.00033	0.00192**	−0.00003	0.000917**	0.00050	0.00328***	0.000691**	0.00517*
	(0.00019)	(0.00037)	(0.00012)	(0.00019)	(0.00024)	(0.00082)	(0.00033)	(0.00045)	(0.00032)	(0.00076)	(0.00034)	(0.00272)
lag3	0.000702***	0.00358***	0.000294***	0.000897***	0.00040	0.00243*	0.00142	0.00139*	0.000868*	0.00528***	0.00142***	0.00456***
	(0.00023)	(0.00069)	(0.00011)	(0.00021)	(0.00040)	(0.00132)	(0.00126)	(0.00078)	(0.00047)	(0.00110)	(0.00050)	(0.00174)
lag4	0.00036	0.00187***	0.000636**	0.000789***	0.00034	0.00459**	0.00034	0.00299	0.00061	0.00441***	0.00141**	0.00201
	(0.00025)	(0.00045)	(0.00027)	(0.00024)	(0.00038)	(0.00218)	(0.00065)	(0.00231)	(0.00055)	(0.00103)	(0.00057)	(0.00126)
lag5	0.00021	0.00193***	0.000488**	0.000689**	0.00207**	0.00193	0.00001	0.00382	0.000762**	0.00762***	0.00127**	0.00487***
	(0.00020)	(0.00050)	(0.00022)	(0.00030)	(0.00093)	(0.00298)	(0.00077)	(0.00249)	(0.00035)	(0.00144)	(0.00060)	(0.00173)
Constant	0.000261***	0.00115***	0.000114***	0.000461***	0.000365***	0.00194***	0.000145*	0.000817***	0.000916***	0.00589***	0.00108***	0.00312***
	(0.00006)	(0.00015)	(0.00004)	(0.00007)	(0.00010)	(0.00024)	(0.00008)	(0.00013)	(0.00011)	(0.00026)	(0.00013)	(0.00031)
Observations	20,622	20,622	20,622	20,622	5,652	5,652	5,652	5,652	22,446	22,446	22,446	22,446
R-squared	0.288	0.337	0.285	0.274	0.272	0.358	0.283	0.321	0.227	0.362	0.279	0.32

	100+ Mbps (vs. <100)				250 Mbps (vs. 100)				Gigabit (vs. 250)			
	Real Estate	Professional	Healthcare	Food & Accom.	Real Estate	Professional	Healthcare	Food & Accom.	Real Estate	Professional	Healthcare	Food & Accom.
lead4	0.00009	0.00011	−0.00004	0.00009	−0.00100	−0.00003	0.00025	−0.00048	−0.000754*	−0.00261***	−0.00814***	−0.00137***
	(0.00013)	(0.00022)	(0.00074)	(0.00012)	(0.00084)	(0.00070)	(0.00250)	(0.00041)	(0.00041)	(0.00074)	(0.00243)	(0.00043)
lead3	−0.00004	0.00009	−0.00024	0.000230*	−0.000727**	−0.00048	−0.00131	−0.000719**	−0.000737*	−0.00127***	−0.00997***	−0.000685*
	(0.00011)	(0.00018)	(0.00054)	(0.00012)	(0.00037)	(0.00046)	(0.00164)	(0.00034)	(0.00039)	(0.00045)	(0.00270)	(0.00041)
lead2	0.00003	0.00001	−0.00007	0.000239**	−0.00085	−0.00027	0.00043	−0.00019	−0.00003	0.00039	−0.00274	−0.00024
	(0.00012)	(0.00016)	(0.00040)	(0.00011)	(0.00062)	(0.00045)	(0.00107)	(0.00044)	(0.00042)	(0.00055)	(0.00217)	(0.00049)
lead1	Reference											
lag0	0.000317***	0.000662***	0.00024	0.000554***	−0.00041	−0.00028	0.00049	0.00034	0.00028	0.00208***	0.00207	0.000775*
	(0.00012)	(0.00024)	(0.00076)	(0.00013)	(0.00060)	(0.00039)	(0.00116)	(0.00038)	(0.00043)	(0.00058)	(0.00213)	(0.00041)
lag1	0.000501***	0.000416**	0.00196***	0.000553***	0.00056	0.00119*	0.00089	0.00151**	0.00106**	0.00247***	0.00314	0.00147***
	(0.00014)	(0.00019)	(0.00045)	(0.00015)	(0.00079)	(0.00065)	(0.00144)	(0.00064)	(0.00047)	(0.00065)	(0.00212)	(0.00043)
lag2	0.000807***	0.000372*	0.00271***	0.000816***	−0.00069	0.00048	0.00409	0.00247**	0.00161**	0.00313***	0.00751**	0.00356***
	(0.00026)	(0.00022)	(0.00067)	(0.00016)	(0.00056)	(0.00055)	(0.00316)	(0.00116)	(0.00076)	(0.00068)	(0.00342)	(0.00079)
lag3	0.000658***	0.000909***	0.00403***	0.00141***	−0.00147	0.00155	−0.00072	0.00137**	0.00146***	0.00547***	0.00865***	0.00367***
	(0.00022)	(0.00030)	(0.00087)	(0.00027)	(0.00092)	(0.00116)	(0.00252)	(0.00062)	(0.00055)	(0.00126)	(0.00280)	(0.00086)
lag4	0.00118***	0.00206***	0.00379***	0.00176***	0.00025	0.00601	−0.00156	0.00585**	0.00154*	0.00524***	0.0145**	0.00526***
	(0.00030)	(0.00048)	(0.00128)	(0.00031)	(0.00150)	(0.00487)	(0.00259)	(0.00269)	(0.00091)	(0.00102)	(0.00614)	(0.00132)
lag5	0.00117***	0.00117***	0.00754**	0.00168***	0.00145	0.00560	0.00498	0.00392	0.00193*	0.00662***	0.0108***	0.00229**
	(0.00029)	(0.00041)	(0.00351)	(0.00038)	(0.00295)	(0.00481)	(0.00625)	(0.00315)	(0.00100)	(0.00170)	(0.00391)	(0.00111)
Constant	0.000350***	0.000650***	0.00229***	0.000270***	0.00121***	0.00132***	0.00447***	0.00114***	0.00257***	0.00416***	0.0164***	0.00303***
	(0.00007)	(0.00009)	(0.00025)	(0.00006)	(0.00030)	(0.00020)	(0.00040)	(0.00015)	(0.00022)	(0.00030)	(0.00155)	(0.00024)
Observations	20,622	20,622	20,622	20,622	5,652	5,652	5,652	5,652	22,446	22,446	22,446	22,446
R-squared	0.248	0.374	0.712	0.262	0.226	0.304	0.559	0.442	0.355	0.538	0.624	0.353

Notes: This table reports event study estimates of the effect of broadband internet speed on business births per 10,000 population at the Census block group level for eight U.S. states, by industry. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed. The lead estimates indicate years prior to treatment. The year prior to receiving treatment is omitted (Lead1). Lag estimates indicate years since treatment. Robust standard errors clustered at the block group level are in parentheses. Asterisks ***, **, and * indicate statistical significance at the 1 %, 5 %, and 10 % levels.

Appendix G. Robustness checks omitting year 2020

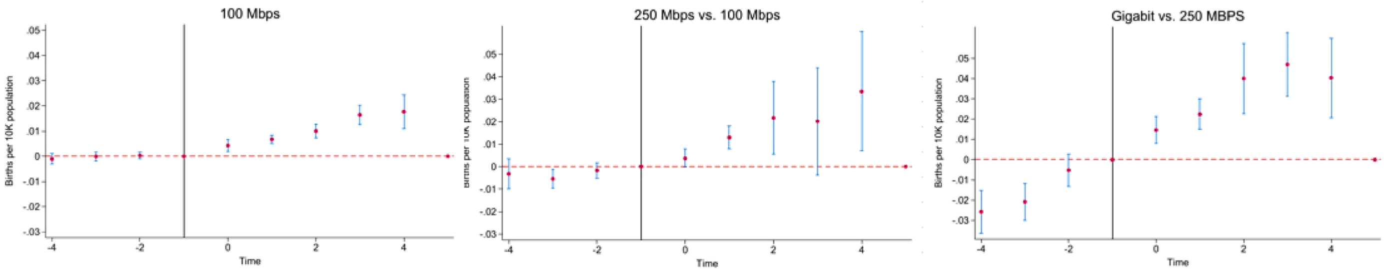

Fig. G.1. Regression of business births per 10,000 population on timing of highspeed broadband access, 2015–2019 (omitting 2020).
Notes: Panels display the event study estimation for the baseline Coarsened Exact Matching (CEM) matched sample. Point estimates are displayed with their 95 % confidence intervals as described in Eq. (2). The baseline (omitted) period is one year prior to treatment in each block group, indicated by the solid vertical line in each plot. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed.

Appendix H. Callaway and Sant'Anna robustness checks

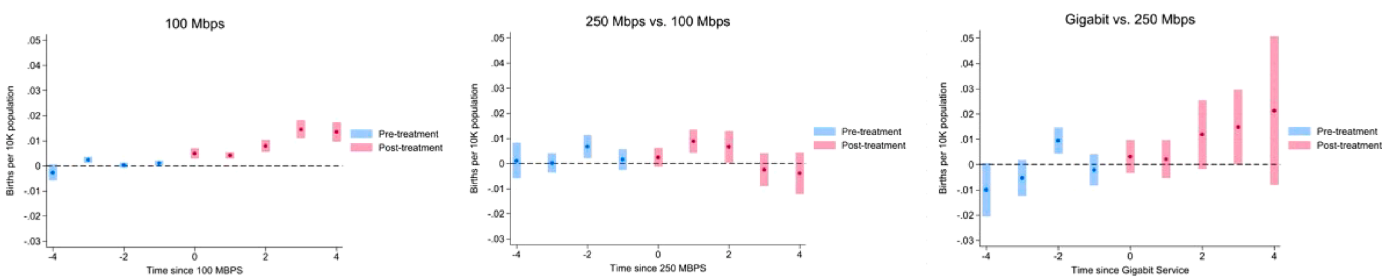

Fig. H.1. Regression of business births per 10,000 population on timing of highspeed broadband access.
Notes: Panels display the event study estimation for the baseline Coarsened Exact Matching (CEM) matched sample, using the Callaway and Sant'Anna (2021) approach. Point estimates are displayed with their 95 % confidence intervals as described in Eq. (2). The baseline (omitted) period is one year prior to treatment in each block group, indicated by the solid vertical line in each plot. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed.

Appendix I. Poisson regression of births (count) on timing of highspeed broadband access, by industry

Table I.1
Poisson regression of births (count) on timing of highspeed broadband access, by industry.

	100+ Mbps (vs. <100)				250 Mbps (vs. 100)				Gigabit (vs. 250)			
	Manufacturing	Retail Trade	Information	Finance	Manufacturing	Retail Trade	Information	Finance	Manufacturing	Retail Trade	Information	Finance
lead4	0.0923	0.105	0.256	0.0323	0.482*	0.133	−0.359	0.448***	0.0806	0.169***	0.0284	−0.131**
	(0.1500)	(0.0711)	(0.2230)	(0.1070)	(0.2720)	(0.1080)	(0.4190)	(0.1710)	(0.1100)	(0.0488)	(0.1270)	(0.0632)
lead3	0.340**	0.171**	0.610***	−0.0533	0.736***	0.0222	0.198	0.237	0.0434	0.317***	0.402***	0.0801
	(0.1540)	(0.0745)	(0.2280)	(0.1150)	(0.2610)	(0.1070)	(0.3350)	(0.1620)	(0.1340)	(0.0596)	(0.1560)	(0.0782)
lead2	0.23	0.169**	0.468*	−0.139	0.628**	−0.000776	0.113	0.315*	−0.00205	0.449***	0.527***	0.0217
	(0.1700)	(0.0798)	(0.2570)	(0.1220)	(0.2820)	(0.1140)	(0.3590)	(0.1710)	(0.1700)	(0.0737)	(0.2000)	(0.0980)
lead1	Reference											
lag0	0.738***	0.159	0.576*	−0.218	0.127	0.118	0.198	0.326**	0.141	0.379***	0.789**	0.0266
	(0.2050)	(0.0991)	(0.3130)	(0.1370)	(0.2000)	(0.0934)	(0.2520)	(0.1450)	(0.2610)	(0.1140)	(0.3070)	(0.1460)
lag1	0.847***	0.189*	0.485	−0.0282	0.485**	0.512***	0.326	0.340**	0.221	0.596***	0.696*	0.355**
	(0.2330)	(0.1120)	(0.3560)	(0.1550)	(0.2230)	(0.1010)	(0.2620)	(0.1570)	(0.3070)	(0.1330)	(0.3630)	(0.1700)
lag2	1.176***	0.229*	0.833**	−0.0553	−0.0311	0.218*	0.225	0.24	0.23	0.694***	1.059**	0.176
	(0.2580)	(0.1240)	(0.3950)	(0.1730)	(0.2900)	(0.1210)	(0.3300)	(0.1670)	(0.3550)	(0.1530)	(0.4170)	(0.1970)
lag3	1.359***	0.513***	1.067**	−0.118	0.33	0.168	0.658	0.0677	0.145	0.842***	1.330***	0.253
	(0.2810)	(0.1340)	(0.4320)	(0.1880)	(0.3830)	(0.1850)	(0.4140)	(0.2010)	(0.4060)	(0.1750)	(0.4730)	(0.2220)
lag4	1.277***	0.402***	0.872*	−0.367*	0.404	0.653**	−0.877	0.257	0.386	0.717***	1.335**	−0.224
	(0.3130)	(0.1470)	(0.4730)	(0.2010)	(0.7350)	(0.3120)	(1.0790)	(0.2970)	(0.4610)	(0.2020)	(0.5390)	(0.2560)
lag5	1.361***	0.437***	0.659	−0.583***	1.245*	0.102	−0.484	0.136	0.15	0.779***	1.159*	0.0805
	(0.3660)	(0.1690)	(0.5300)	(0.2220)	(0.6600)	(0.5530)	(1.0470)	(0.4060)	(0.5970)	(0.2560)	(0.6760)	(0.3140)
Observations	8,286	16,890	5,028	10,122	2,226	4,746	1,650	3,108	8,520	18,498	8,118	13,566
	100+ Mbps (vs. <100)				250 Mbps (vs. 100)				Gigabit (vs. 250)			

(*continued on next page*)

Table I.1 (*continued*)

	100+ Mbps (vs. <100)				250 Mbps (vs. 100)				Gigabit (vs. 250)			
	Real Estate	Professional	Healthcare	Food & Accom	Real Estate	Professional	Healthcare	Food & Accom	Real Estate	Professional	Healthcare	Food & Accom
	Real Estate	Professional	Healthcare	Food & Accom	Real Estate	Professional	Healthcare	Food & Accom	Real Estate	Professional	Healthcare	Food & Accom
lead4	0.0553	−0.201*	0.177***	−0.176	−0.186	0.255*	0.305***	−0.436**	0.0155	0.166***	0.120***	0.136*
	(0.1540)	(0.1090)	(0.0623)	(0.1080)	(0.2220)	(0.1380)	(0.0870)	(0.2060)	(0.0908)	(0.0629)	(0.0324)	(0.0721)
lead3	0.0125	−0.210*	0.185***	−0.0946	−0.25	0.0439	0.0725	−0.0995	−0.111	0.347***	0.0756*	0.204**
	(0.1590)	(0.1130)	(0.0631)	(0.1120)	(0.1740)	(0.1220)	(0.0770)	(0.1350)	(0.1120)	(0.0784)	(0.0406)	(0.0855)
lead2	0.336**	−0.196*	0.0505	−0.0522	−0.233	0.116	0.0873	−0.00174	−0.09	0.498***	0.217***	0.291***
	(0.1610)	(0.1190)	(0.0694)	(0.1160)	(0.1470)	(0.1050)	(0.0656)	(0.1090)	(0.1420)	(0.0997)	(0.0523)	(0.1040)
lead1	Reference											
lag0	0.366*	−0.231	0.116	0.13	0.19	−0.0793	0.00319	0.235*	−0.125	0.706***	0.273***	0.152
	(0.1970)	(0.1410)	(0.0839)	(0.1390)	(0.1800)	(0.1230)	(0.0719)	(0.1350)	(0.2170)	(0.1530)	(0.0794)	(0.1590)
lag1	0.398*	−0.284*	0.169*	0.146	0.166	0.132	−0.0641	0.748***	−0.0877	0.989***	0.391***	0.233
	(0.2200)	(0.1600)	(0.0938)	(0.1570)	(0.2010)	(0.1330)	(0.0890)	(0.1400)	(0.2550)	(0.1800)	(0.0933)	(0.1850)
lag2	0.625**	−0.326*	0.152	0.328*	0.371*	0.00517	−0.02	0.707***	−0.324	1.053***	0.380***	0.374*
	(0.2430)	(0.1790)	(0.1040)	(0.1740)	(0.2190)	(0.1470)	(0.1030)	(0.1670)	(0.2950)	(0.2080)	(0.1080)	(0.2120)
lag3	0.676**	−0.387**	0.244**	0.556***	0.413	0.0735	−0.261*	0.426	−0.245	1.304***	0.461***	0.327
	(0.2640)	(0.1940)	(0.1130)	(0.1880)	(0.3560)	(0.2480)	(0.1400)	(0.2630)	(0.3340)	(0.2360)	(0.1220)	(0.2410)
lag4	0.591**	−0.305	0.161	0.698***	0.639	0.0701	−0.289	1.101**	−0.214	1.376***	0.452***	0.171
	(0.2880)	(0.2090)	(0.1230)	(0.2040)	(0.4160)	(0.4470)	(0.1830)	(0.5100)	(0.3790)	(0.2670)	(0.1390)	(0.2750)
lag5	0.998***	−0.322	0.235*	0.866***	1.486***	0.181	−0.0575	0.965**	−0.306	1.487***	0.502***	0.2
	(0.3200)	(0.2330)	(0.1360)	(0.2310)	(0.4700)	(0.4320)	(0.2690)	(0.4530)	(0.4610)	(0.3200)	(0.1700)	(0.3350)
Observations	9,222	12,294	12,882	11,598	2,604	3,486	3,804	3,438	11,490	15,150	16,446	13,872

Notes: This table reports event study Poisson regression estimates of the effect of broadband internet speed on business births at the Census block group level for eight U. S. states, by industry. Treatment is defined as greater than or equal to 50 % of establishments in a block group having access to 100+ Mbps, 250 Mbps, or gigabit (1GB = 1000 Mbps) broadband download speed. The lead estimates indicate years prior to treatment. The year prior to receiving treatment is omitted (Lead1). Lag estimates indicate years since treatment. Robust standard errors clustered at the block group level are in parentheses. Asterisks ***, **, and * indicate statistical significance at the 1 %, 5 %, and 10 % levels. Block groups with all zero outcomes are omitted from the analysis.

References

Acs, Z., 2006. How is Entrepreneurship Good for Economic Growth? Innovations 1 (1), 97–107.

Alderete, M.V., 2017. Mobile broadband: a key enabling technology for entrepreneurship? J. Small Bus. Manage. 55 (2), 254–269.

Ashenfelter, O., Card, D, 1985. Using the longitudinal structure of earnings to estimate the effect of training programs. Rev. Econ. Stat. 67 (4), 648–660.

Atasoy, H., 2013. The effects of broadband internet expansion on labor market outcomes. ILR Rev. 66 (2), 315–345.

Audretsch, D.B., Heger, D., Veith, T., 2015. Infrastructure and entrepreneurship. Small Bus. Econ. 44, 219–230.

Austin, P.C., 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat. Med. 28 (25), 3083–3107.

Ayar Labs, 2022. How Giant AI Workloads and the Looming 'Bandwidth Wall' Are Impacting System Architectures. HPCWire. October 24, 2022. Available online: https://www.hpcwire.com/2022/10/24/how-giant-ai-workloads-and-the-looming-bandwidth-wall-are-impacting-system-architectures/.

Bertoni, D., Curzi, D., Aletti, G., Olper, A., 2020. Estimating the effects of agri-environmental measures using difference-in-difference coarsened exact matching. Food Policy 90.

Bertschek, I., Briglauer, W., Hüschelrath, K., Kauf, B., Niebel, T., 2015. The economic impacts of broadband internet: a survey. Rev. Netw. Econ. 14 (4), 201–227.

Busby, J., Tanberk, J., 2020. FCC Reports Broadband Unavailable to 21.3 Million Americans, BroadbandNow Study Indicates 42 Million Do Not Have Access. BroadbandNow Research. Available online: https://broadbandnow.com/research/fcc-underestimates-unserved-by-50-percent.

Busby, J., Tanberk, J., Cooper, T, 2021. BroadbandNow Estimates Availability for all 50 States; Confirms that More than 42 Million Americans Do Not Have Access to Broadband. BroadbandNow Research. Available online: https://broadbandnow.com/research/fcc-broadband-overreporting-by-state.

Callaway, B., Sant'Anna, P., 2021. Difference-in-differences with multiple time periods. J. Econom. 225 (2), 200–230.

Clarke, D. and Schythe, K. (2020). Implementing the Panel Event Study. IZA Discussion Paper No. 13524. Available online: https://www.iza.org/publications/dp/13524/implementing-the-panel-event-study.

Conley, K., Whitacre, B.E., 2020. Home is where the internet is? High-speed internet's impact on rural housing values. Int. Reg. Sci. Rev. 43 (5), 501–530.

Conlow, M., 2023. Congratulations to all the states, NTIA, and FCC on the allocation of broadband funds…and new spreadsheet with the allocations. Mike Conlow Substack. Available online: https://mikeconlow.substack.com/p/congratulations-to-all-the-states.

Conroy, T., Low, S., 2022. Entrepreneurship, broadband, and gender: evidence from establishment births in rural America. Int. Reg. Sci. Rev. 45 (1), 3–35.

Crandall, R.W., Lehr, W., Litan, R., 2007. The Effects of Broadband Deployment On Output and employment: A cross-Sectional Analysis of U.S. Data. Brookings Institution Press, Washington, D.C.. Available online: https://www.brookings.edu/wp-content/uploads/2016/06/06labor_crandall.pdf

Czernich, N., 2014. Does broadband internet reduce the unemployment rate? Evidence for Germany. Inf. Econ. Policy 29, 32–45.

Data Axle (2022). ``Our Data." Data Axle Reference Solutions. https://www.data-axle.com/our-data/business-data/.

de Chaisemartin, C., D'Haultfoeuille, X, 2022. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. Econom. J. 00, 1–30.

Decker, R. and Haltiwanger, J. (2022). Business entry and exit in the COVID-19 pandemic: a preliminary look at official data. FEDS Notes, May 6. Available online: https://www.federalreserve.gov/econres/notes/feds-notes/business-entry-and-exit-in-the-covid-19-pandemic-a-preliminary-look-at-official-data-20220506.html.

Deller, S., Whitacre, B.E., 2019. Broadband's relationship to rural housing values. Pap. Reg. Sci. 98 (5), 2135–2156.

Deller, S., Whitacre, B.E., Conroy, T., 2022. Rural broadband speeds and business startup rates. Am. J. Agric. Econ. 104, 999–1025.

Deloitte. (2022). "2022 Connectivity and Mobile Trends Survey, 3rd ed." Available online: https://www2.deloitte.com/us/en/insights/industry/telecommunications/connectivity-mobile-trends-survey.html.

Dharmasankar, S., Yoo, H., 2023. Assessing the main and spillover effects of Seattle's minimum wage on establishment decisions. Reg. Sci. Urban Econ. 99.

Duvivier, C., Bussière, C., 2022. The contingent nature of broadband as an engine for business startups in rural areas. J. Reg. Sci. 62 (5), 1329–1357.

Duvivier, C., Cazou, E., Truchet-Aznar, S., Brunelle, C., Dubé, J., 2021. When, where, and for what industries does broadband foster establishment births? Pap. Reg. Sci. 100 (6), 1377–1401.

Fabling, R., Grimes, A., 2021. Picking up speed: does ultrafast broadband increase firm productivity? Inf. Econ. Policy 57.

Falk, M., Hagsten, E., 2021. Impact of high-speed broadband access on local establishment dynamics. Telecomm. Policy 45 (4).

Federal Communications Commission (FCC). (2015). Fifth Measuring Broadband America Fixed Broadband Report. February 4, 2015. Available online: https://www.fcc.gov/reports-research/reports/measuring-broadband-america/measuring-broadband-america-2015.

Federal Communications Commission (FCC). (2022). Fixed Broadband Deployment Data from FCC Form 477. Retrieved from: https://www.fcc.gov/general/broadband-deployment-data-fcc-form-477.

Federal Communications Commission (FCC). (2022). Information for filers." Available online: https://www.fcc.gov/BroadbandData/filers.

Federal Communications Commission (FCC). (2022). Public Notice. "Broadband data task force announces the start of the Broadband Serviceable Location Fabric bulk challenge process." September 1, 2022. Available online: https://docs.fcc.gov/public/attachments/DA-22-913A1.pdf.

Federal Communications Commission (FCC). (2024). FCC Increases Broadband Speed Benchmark. Available online: https://docs.fcc.gov/public/attachments/DOC-401205A1.pdf.
Ford, G.S., 2018. Is faster better? Quantifying the relationship between broadband speed and economic growth. Telecomm. Policy 42 (9), 766–777.
Ford, G.S. (2022). Overstating broadband availability: an Assessment of the “All-in” assumption for FCC 477 Data. Phoenix Center for Advanced Legal and Economic Public Policy Studies Perspectives. Available online: https://www.phoenix-center.org/perspectives/Perspective22-04Final.pdf.
Gallardo, R., Whitacre, B.E., 2024. An unexpected digital divide? a look at internet speeds and socioeconomic groups. Telecommun. Policy 48 (6), 102777. https://doi.org/10.1016/j.telpol.2024.102777.
Galloway, L., Sanders, J., Deakins, D, 2011. Rural small firms’ use of the internet: from global to local. J. Rural Stud. 27 (3), 254–262.
Goepel, T. (2018). “Artificial Intelligence and Machine Learning Require a Better Network.” TechNative. October 12, 2018. Available online: https://technative.io/artificial-intelligence-and-machine-learning-require-a-better-network/.
Han, L., Wojan, T.R., Goetz, S.J., 2023. Experimenting in the cloud: the digital divide’s impact on innovation. Telecomm. Policy. In Press.
Hasbi, M., Bohlin, E., 2022. Impact of broadband quality on median income and unemployment: evidence from Sweden. Telematics Inf. 66.
Hasbi, M., 2020. Impact of very high-speed broadband on company creation and entrepreneurship: empirical evidence. Telecomm. Policy 44, 101873.
Iacus, S.M., King, G., Porro, G., 2009. CEM: coarsened exact matching software. J. Stat. Softw. 30 (9). Available online: http://gking.harvard.edu/cem.
Iacus, S.M., King, G., Porro, G., 2011. Multivariate matching methods that are monotonic imbalance bounding. J. Am. Stat. Assoc. 106 (493), 345–361.
Iacus, S.M., King, G., Porro, G., 2012. Causal inference without balance checking: coarsened exact matching. Polit. Anal. 20 (1), 1–24.
Kandilov, I., Renkow, M., 2010. Infrastructure investment and rural economic development: an evaluation of USDA’s broadband loan program. Growth Change 41 (2), 165–191.
Karunathilake, E., Le, A., Heo, S., Chung, Y., Mansoor, S., 2023. The path to smart farming: innovations and opportunities in precision agriculture. Agriculture 13, 1593.
Kim, Y., Orazem, P., 2016. Broadband internet and new firm location decisions in rural areas. Am. J. Agric. Econ. 99 (1), 285–302.
King, G., Nielsen, R., 2019. Why propensity scores should not be used for matching. Polit. Anal. 27 (4), 435–454.
Kolko, J., 2012. Broadband and local growth. J. Urban Econ. 71 (1), 100–113.
Koutroumpis, P., 2019. The economic impact of broadband: evidence from OECD countries. Technol. Forecast. Soc. Change 148.
Lobo, B.J., Alam, M.R., Whitacre, B.E., 2020. Broadband speed and unemployment rates: data and measurement issues. Telecomm. Policy 44 (1).
Mack, E., Loveridge, S., Keene, T., Mann, J., 2023. A review of the literature about broadband internet connections and rural development (1995–2022). Int. Reg. Sci. Rev. 0 (0) https://doi.org/10.1177/01600176231202457.
Mahasuweerachai, P., Whitacre, B.E., Shideler, D.W., 2010. Does broadband access impact migration in America? Examining differences between rural and urban areas. Rev. Reg. Stud. 40 (1), 5–26.
Major, D., Teixeira, R., Mayer, J., 2020. No WAN’s land: mapping U.S. broadband coverage with millions of address queries to ISPs. In: Proceedings of the ACM internet measurement conference, pp. 393–419. October 2020.
Maude, H., 2020. Impact of very high-speed broadband on company creation and entrepreneurship: empirical evidence. Telecomm. Policy 44 (3).
Meltzer, R., Ellen, I., Li, X., 2021. Localized commercial effects from natural disasters: the Case of Hurricane Sandy and New York City. Reg. Sci. Urban Econ. 86, 103608.
National Telecommunications and Information Administration (NTIA). (2023). Biden-Harris Administration Announces State Allocations for $42.45 Billion High-Speed Internet Grant Program. Available online: https://www.ntia.gov/press-release/2023/biden-harris-administration-announces-state-allocations-4245-billion-high-speed.
Nichols, A., 2010. Regression for nonnegative skewed dependent variables. In: Stata Conference presentation. Available online: https://www.stata.com/meeting/boston10/boston10_nichols.pdf.
O’Shea, D., 2022. FCC’s Rosenworcel Proposes ‘100/20’ as New Broadband Minimum Requirement. Fierce Telecom. Available online: https://www.fiercetelecom.com/broadband/fccs-rosenworcel-proposes-10020-new-broadband-minimum-requirement.
Ripollone, J., Huybrechts, K., Rothman, K., Ferguson, R., Franklin, J, 2020. Evaluating the utility of coarsened exact matching for pharmacoepidemiology using real and simulated claims data. Am. J. Epidemiol. 189 (6), 613–622.
Rupasingha, A., Pender, J., Williams, R., Goldstein, J., Nair, D., 2023. Place-based subsidies and employment growth in rural America: evidence from the broadband initiatives programme. Pap. Reg. Sci. early view.
Sarachuk, K., Missler-Behr, M., 2020. Is Ultra-broadband enough? The relationship between high-speed internet and entrepreneurship in Brandenburg. Int. J. Technol. 11 (6), 1103–1114.
Shaikh, T., Rasool, T., Lone, F., 2022. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. Comput. Electron. Agric. 198, 107119.
Shideler, D., Badasyan, N., 2012. Broadband impact on small business growth in Kentucky. J. Small Bus. Enterprise Dev. 19 (4), 589–606.
Stenberg, P., 2009. Broadband Internet’s Value For Rural America. Department of Agriculture, Washington DC: U.S. Economic Research Report.
Sternberg, R., Wennekers, S., 2005. Determinants and effects of new business creation using global entrepreneurship monitor data. Small Bus. Econ. 24, 193–203.
Sun, L., Abraham, S., 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. J. Econom. 225 (2), 175–199.
Supan, J., 2021. Advertised vs. Actual Internet Speeds – Millions of Americans Aren’t Getting What They Pay For, online. Allconnect. https://www.allconnect.com/blog/advertised-vs-actual-internet-speeds.
Taylor, D.E., Bell, A., Saherwala, A., 2022. Understanding food access in Flint: an analysis of racial and socioeconomic disparities. Am. Behav. Sci. In Press.
Theodos, B., Stacy, C., Teles, D., Davis, C., Hariharan, A., 2022. Place-based investment and neighborhood change: the impacts of New Markets Tax Credits on jobs, poverty, and neighborhood composition. J. Reg. Sci. 62 (4), 1092–1121.
United States Census Bureau. (2022). “Glossary.” Available online: https://www.census.gov/programs-surveys/geography/about/glossary.html#par_textimage_4.
United States Department of Transportation (U.S. DOT), 2022. Roadside Telecommunications (RS-TC) Fiber Optic Cable Installation. Intelligent Transportation Systems Deployment Evaluation. Available online. https://www.itskrs.its.dot.gov/decision-support/cost-plot/roadside-telecommunications-rs-tc-fiber-optic-cable-installation.
Urbano, D., Aparicio, S., Audretsch, D., 2019. Twenty-five years of research on institutions, entrepreneurship, and economic growth: what has been learned? Small Bus. Econ. 53, 21–49.
Van Leuven, A.J., 2021. The impact of Main Street revitalization on the economic vitality of small-town business districts. Econ. Dev. Q. 36 (3), 193–207.
Whipple, T., 2023. Fiber Technology will Dominate BEAD Deployment, Agree Panelists. Broadband Breakfast. Available online: https://broadbandbreakfast.com/2023/05/fiber-technology-will-dominate-bead-deployment-agree-panelists-at-wireless-conference/.
Whitacre, B.E., Biedny, C., 2022. A preview of the broadband fabric: opportunities and issues for researchers and policymakers. Telecomm. Policy 46 (3).
Whitacre, B.E., Gallardo, R., Strover, S., 2014a. Does rural broadband impact jobs and income? Evidence from spatial and first-differenced regressions. Ann. Reg. Sci. 53 (3), 649–670.
Whitacre, B.E., Gallardo, R., Strover, S., 2014b. Broadband’s contribution to economic growth in rural areas: moving towards a causal relationship. Telecomm. Policy 38 (11), 1011–1023.
Whitacre, B.E., 2011. Do higher broadband adoption rates mean lower tax collections from local retail sales? Implications of e-commerce in rural areas of the U.S. Reg. Sci. Policy Pract. 3 (2), 71–85.
Wolf, D., Irwin, N., 2023. Is it really bridging the gap? Fiber internet’s impact on housing values and home-buyer demographics. J. Reg. Sci. 64, 238–271. https://doi.org/10.1111/jors.12670.
Wooldridge, J. (2020). Statalist forum on Poisson /Negative Binomial regressions. Available online: https://www.statalist.org/forums/forum/general-stata-discussion/general/1587040-why-do-poisson-and-negative-binomial-regressions-yield-the-same-result.