Internet's Important Involvement In Information Industry Integration In Idaho, Iowa, Illinois, Indiana (and others):

.

How the emerging internet affected the economic geography of the information industry *

Undergraduate Honors Thesis

Keming (Alex) Gao

May 16, 2020

1 Abstract

I explore the effect of early competitive advantages in internet communication technology (ICT) and ICT-adjacent human capital on the economic geography of the information industry. The paper uses assignment of NSFNET Regional Network funding to certain MSAs as a signifier of infrastructure and human capital development related to networking. This "treatment" of infrastructure/human capital, despite occurring in 1985, lay dormant for almost a decade - businesses were fully banned from accessing both the national and the regional networks, there were no other national networks, and the "website" didn't even exist yet. However, from 1994-95, businesses were suddenly given access, the World Wide Web came into being, and the "treatment" suddenly activated, conferring a temporary competitive advantage to businesses in those areas. I perform two difference-in-difference regressions (pooled, fixed effects) on different industry outcomes, an event study regression to discuss parallel trends, and a regression to correlate distance from the network node with centralization outcomes. I analyze the results from an urban economics standpoint. Ultimately, I argue that increased accessibility to ICT-adjacent resources in the early days had some long term effects on the centralization of the rapidly transforming information industry, even though distance is no longer a real factor for ICT applications. I also discover an interesting avenue for future research (the negative supply shock instigated by the dot-com boom, in which the treated MSAs recovered faster/were totally unaffected).

*I would like to thank all my professors (advisor David Card, James Sallee, Patrick Kline, Barry Eichengreen, Cecile Gaubert, Christina and David Romer, Demian Pouzo, and Raymond Hawkins) and graduate students/post-docs Luna Huang, Hannah Druckenmiller, Nick Li, and Sergio Castellanos for facilitating and supporting my academic interests. Thank you all for devoting your time to educating me.

2 Introduction

I introduce my motivation, the history and context surrounding this topic, the research question, and my methodology. The rest of the paper overviews the existing research in this area, explains where the data is from and how it is processed, describes the identification strategy, and discusses results.

2.1 Motivation

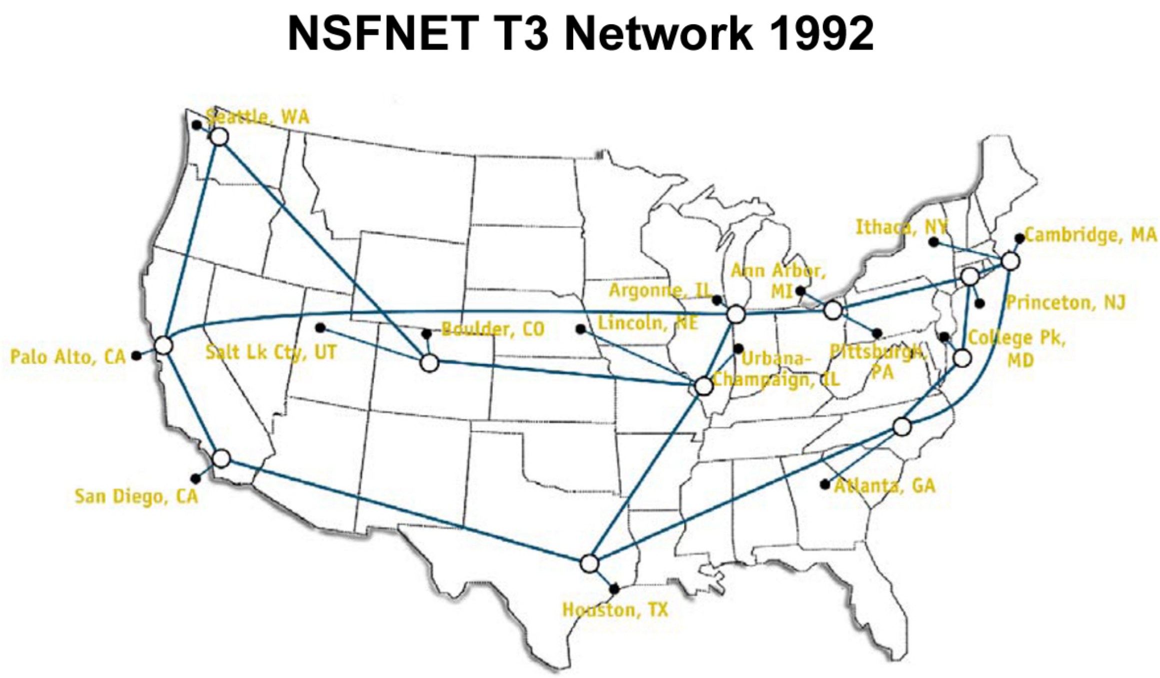
Internet communication technology (ICT) has undoubtedly had a massive effect on the economy. However, there is still little understanding of how existing technology has impacted the economy of today. ICT has only been widespread for a little over 20 years - the World Wide Web was invented only in 1991, and it was only in April 1995 that full commercial use of the internet was approved. The technology was still seen as "emerging" for a long time, so much so that Paul Krugman said in 1998 that the Internet's economic impact on the world would be "no greater than the fax machine's." The point is, ICT's relationship to the development of our modern economy is still poorly understood - in part because it happened so quickly, in part because it may be fundamentally different from traditional physical goods, in part because the technology itself (and the cultural context) is changing so rapidly. However, it is extremely relevant - for instance, recent decisions by companies like Twitter to allow permanent telework has driven speculations about how the San Francisco Bay Area's housing and labor markets may fundamentally change.

I am motivated to take a historical view of the relationship between ICT and the economy. At face value, the situation in the past has little relation to today - the technology itself and the economic context is so incredibly different. But the early technology may have had an institutional impact that reverberates to this day. In the same way that river networks impacted the development of early American cities, even though cities can now easily locate much further away from local water sources, it is possible that early centers of the information industry were affected by access to ICT infrastructure and human resources, and agglomeration continued around those urban centers even after the initial competitive advantages went away.

Thus, I look to the early 1990s - an era when commercial access to internet was not allowed, the first network infrastructures were being built, and the "website" didn't even exist.

2.2 NSFNET/Internet History and Context

In 1985, the NSF proposed the construction of five supercomputing centers at universities in Ithaca, Boulder, San Diego, Princeton, and Pittsburgh. To connect these centers, they began building the NSFNET and funding networking initiatives at these universities. Over the next three years, they also funded the creation, expansion, and connection of regional networks in 11 MSAs, called the NSFNET Regional Networks [13]. The NSFNET was the fastest, highest capacity, and largest network in the United States - it was the **only** national network accessible to the public - and it **banned all commercial use**, only allowing for academic research [5]. Though businesses could theoretically access private and more localized networks, there was little incentive to - use cases were extremely limited, consumers had almost no access, and there was no such thing as the Internet or the World Wide Web yet.



Beginning in 1994, commercial restrictions on private access were lifted across all the NSFNET - and consequently, the World Wide Web, which was just invented a year before - for the first time [6]. The infrastructure and human capital that the NSFNET had invested in suddenly had massive commercial potential.

Some businesses and consumers very suddenly had access to mature infrastructure and people who were familiar with the potential of the Internet. Those located at the initial NSFNET sites could learn about the technology faster, and connect faster without waiting for commercial ISPs to build out new infrastructure. The competitive advantage of being close to these NSFNET MSAs decreased very quickly - access and knowledge exploded across the country all at once, and by the end of 1996, Internet access was already a truly national phenomenon.

2.3 Research question and hypothesis

The situation can be highly simplified to having three time periods. In period one (pre-1994), businesses don't know what the internet even is, and couldn't access it even if they wanted to. In period two (1994-1995), some businesses suddenly have high accessibility to internet infrastructure, and to human beings who have an unique understanding of ICT's capabilities. In period 3 (1996-onwards), most businesses have internet access if they want it, and the technology is layman-accessible, meaning most of the special competitive advantage in period 2 is gone.

Period two is of utmost importance. In that narrow time frame, the "treatment" of human capital and infrastructure was activated, and certain competitive advantages were suddenly conferred onto the information industry in these MSAs. These temporary advantages could have snowballed into agglomeration effects that reverberated to future commercial developments. These effects aren't just about industry growth and centralization - they could also include the reshaping of urban economies such that they are more "defined" by certain industries.

Decomposing the precise effect of infrastructure as opposed to human capital is hard, but the exact effect sizes are not of particular policy interest - this specific context will surely never be replicable again, and the specific effect sizes will never really be applicable to the future.

The relevant policy question is whether these short-lived competitive advantages in ICT/ICT-adjacent human capital had long lasting effects that entrenched themselves, or whether the rapid spread of that knowledge/infrastructure had a "globalizing" effect that outweighed the advantages conferred by "initial conditions". In this specific historical context, the question is, "did the areas that suddenly had a competitive advantage in ICT in 1994 experience long-term higher growth in the information industry? Additionally, did they reshape around the information industry"

I hypothesize increased information industry centralization and growth in certain relevant sectors (e.g. internet publishing, but not necessarily music recording), in the MSAs selected by the NSFNET to receive funding for regional networks. Basically, when the government empowers certain winners, even temporarily, it has lasting effects when it comes to ICT and consequently the economic geography of industries which rely on it.

I further hypothesize that the centralization effect will vary with distance to the node, implying that infrastructure access in the early stages is a contributor to this effect (this is similar to the urban economics interpretation of how increasing transport costs lead to centralization of an industry around a particular resource).

2.4 Methodology

My methodology has 4 steps.

First, I establish at a base level whether the information industry centralized or not in the nation. If the information industry was centralizing, my goal would be to find hotspots where it was centralizing the most. If it was decentralizing, I would try to find areas where the industry was spreading out towards, which would require a different strategy.

Second, I choose outcome variables. I define a way to quantify 'centralization' at a local level, since there are varied ways to do this in the existing literature with no authoritative answer. The index I define encodes a measure of both industry centralization into specific MSAs and also industry primacy in an MSA compared to other industries.

Third, I set up for my identification strategy by discussing the "treatment" of NSFNET node assignment. I make qualitative and historical arguments against reverse causation (industry qualities affecting NSFNET node assignment). I also discuss that even though the NSFNET nodes may have been assigned on unobservables such as "networking research potential", those unobservables did not actually "activate" to affect industry until after 1994, and thus the "initial conditions" of our industries of interest are not necessarily correlated with the assignment.

Fourth, I explore some identification strategies. I settle on exploring 8 relevant NAICS industry sectors. For outcome variables, I select 3 (average pay, employment levels, and num of establishments) and also explore a comparative representation of these variables that encodes different understandings of what it means to "centralize". I explore 2 differences-in-differences (DID) specifications - a pooled two-period model and a DID with state-time fixed effects. I finally discuss the parallel trends assumption by comparing effect sizes in an event study.

Fifth, I make some extensions to the exploration above that may be interesting, but ultimately not very rigorous. I use differences in the treatment group in an attempt to understand the relative advantage of infrastructure access versus human capital. I also take results from part 4 and make spatial correlations about whether distance from the central node is a relevant factor in the development of those industries, connecting back to certain urban economics concepts.

I finish with an exploration of the results and a discussion of limitations and future work.

3 Literature Review

In this section, I go more in depth into NSFNET/Internet history, which provides crucial context for the validity of my identification strategy. I further explain two opposing concepts called the 'network city' and the 'global city', overview the theoretical and empirical arguments for both sides, and discuss their applications to this field. Then, I will survey some of the holes in the existing empirical research. Finally, I will explain how my paper fits into the existing literature and debate.

3.1 NSFNET/Internet history

The NSFNET Backbone and NSFNET Regional Networks were first conceived in 1985. By 1990, all the regional sites had connected to the Backbone, with no commercial usage allowed. By June 1992, some commercial access was allowed for experimental purposes - but under the list of "Unacceptable Uses" in the terms of agreement was "use for for-profit activities" and "use for private or personal business." [6]

All this high speed infrastructure and regional networking capability was being created in these areas [15], but no businesses were benefiting from it. Consumers (if they weren't researchers/students) didn't benefit either. Even though they could have built access to localized private networks, there wasn't much to do with it - to provide some perspective, at the beginning of 1993, there were only 26 websites (none commercial) and using the web still required royalty payments to CERN. [5]

1994 was a watershed year. CERN made the technology free. The first real browser, which could handle commercial applications like payments, was released. The central NSFNET network was sold to commercial ISPs, and commercial restrictions on the NSFNET regional networks were lifted. Management and routing of existing regional networks were distributed to commercial ISPs, who could now provide easy access to consumers. [14]

Some were located very close to these central regional networks - others were further away. Some regional networks were still isolated rather than connected to this national infrastructure. Thus, there was a temporary competitive advantage conferred by location. Being near the NSFNET Regional Network MSAs meant near-immediate access to a national network (that had international connections as well!), as well as access to the local expertise that potentially understood the power of the nascent internet better.

This competitive advantage would erode very quickly. By 1996, there were 257,000 websites (https://www.internetlivestats.com/total-number-of-websites/). That's almost a 1 million percent change in 2 years. Internet access and internet knowledge spread across the country incredibly rapidly, and in the span of two short years, we were in the beginning of the dot com boom.

3.2 Theories of urban centralization

I first explore the main schools of thought surrounding ICT's effect on urban areas - the 'network city' and the 'global city' concepts.

The 'network city' concept argues that due to ICT, cities will lose their status as central hubs of economic activity. In the world of the 'network city', it doesn't matter if an area has a competitive ICT advantage or whether it lasts, since ICT fundamentally decentralizes industries.

The theoretical backing for this involves two forces: transportation costs, and 'face-time' [10, 16]. These forces, in the urban economics literature, are regarded as the dominant forces that shape the geography of cities as a **general theory**, not just for a specific firm or situation. Firms tend locate themselves on the basis of transportation costs – if it is cheaper to ship inputs a shorter distance, they will locate as close as possible to cities to save money until land price equilibrium is reached. Human beings and human-capital tends to highly value 'face-time'; the idea is that being able to see your business partners face to face has major economic benefits and enables trust, negotiations, and cooperation.

In the case of ICT, it becomes easier to coordinate logistics and shipping for traditional physical inputs like fabrics or metals, driving transportation costs down and allowing firms to locate away from a city center. Furthermore, with new access to ICT, firms can now quite literally FaceTime even if they are not located close to each other, further allowing firms to decentralize and spread out their operations. With ICT, physical proximity is no longer a communication barrier, and different components of businesses need not be limited by location.

On an empirical level, there are two papers which support the idea that decentralization occurs. The first studies an area in Germany and finds that there is an overall trend towards decentralization and firms spreading their operations further out [3]. Another approach analyzes domain name registrations and find that they are increasingly spread out, away from traditional business centers [18].

The 'global city' is direct opposition to the idea of the network city, and also has a theoretical and economic basis. In this conception, ICT functions very much like traditional technologies, and the agglomeration effects that apply for traditional resources and technologies still apply here.

Once again, the arguments center on the dominant forces in urban economics (transportation costs and face-time). Transportation costs ultimately may not change much, as information costs were near negligible to begin with. Additionally, while ICT would drive transportation costs down, it would not be enough to cause structural economic changes – at best, we would see more spread out cities rather than wholesale decentralization [12]. Additionally, the real value of face-time lies in creating trust and understanding beyond mere words – otherwise, even handwritten mail could stand in for face-time. ICT enables digital conversations, but does not enable the 'digital handshakes' and trust that are increasingly necessary in a more complex economy [12].

Another theoretical viewpoint treats the Internet as a good, just like any other, and whether the Internet is a substitute for or a complement to cities. One perspective holds that the Internet does not substitute for cities. Since the Internet is a facilitator for the sale and purchase of goods, the question that should be asked is whether the goods being sold are still local in nature or not – and the empirical evidence indicates that those goods remain local (Sinai 2004). Analyses of specific industries like foreign exchange also support this conclusion, indicating that within countries, more financial activity moves towards traditional financial centers, even as the financial activity becomes more spread out globally [8].

When applied to the NSFNET "treatment" that infused ICT related infrastructure and human-capital, we can see similar themes come into play. After the treatment was made accessible for businesses in 1994, businesses with local access to existing infrastructure surely faced lower costs for ICT access, and had an easier time getting face-time with people who were ICT-savvy as well - but only 1-2 years later, the technology had already spread so incredibly rapidly that those advantages were almost certainly reduced. Thus, both centralizing and decentralizing effects were acting, but on different time scales - there was an initial centralizing push, but a longer term decentralizing force with the spread of technology.

3.3 Gaps in the existing research

The most important gap in the existing research is that there is **no comparative approach** being done. There is almost no way to compare the development of one economy with ICT and another without it, for the simple reason that in modern times, almost **everyone** has ICT access, and existing disparities in ICT access are surely correlated with broader economic factors that confound the overall economic situation.

This paper attempts a faux-comparative approach by leveraging the temporary variation in ICT use/access induced by the NSFNET opening. This variation may have been disconnected from the broader economic context, given that the initial NSFNET nodes were assigned on the basis of logistical and scientific reasons rather than economic ones (this is an assertion we explore more in the methodology section).

Furthermore, the 'network city' and 'global city' analyses suffer from 3 problems: (1) it is difficult to correlate economic activity to the vague notion of 'technology', (2) they lack an indicator for economic activity that can be meaningfully tied to ICT, and (3) they do not use an industry-level approach. The Franz-Josef Bade analysis of Germany has no statistical method used to correlate ICT and the finding of decentralization in any way. The only argument made is one of coinciding time frames.

Additionally, Townsend's use of domain name registrations as an economic indicator may not be valid because domain name purchases are not good indicators of meaningful employment or economic output. Finally, Sinai's connection of consumer usage of ICT to what kinds of goods are bought does not take an industry level approach. Perhaps the consumers are purchasing more local goods, but the industry supplying those local goods is now in many more cities, helping to grow/manufacture those local goods.

Overall, in my methodology, I find an exogenous indicator of "ICT access", adopt an identification strategy that meaningfully ties an economic outcome to ICT, and find data for a broad industry-level approach. Most importantly, this is a comparative analysis between a group that received early ICT access and another that did not.

4 Data

4.1 NSFNET Data

The NSF has a record of different stages of the NSFNET network over time. I am interested in the NSFNET as it existed in 1995, right before it was opened up to commercial involvement. The data indicates that nodes existed in these MSAs: Palo Alto, Seattle, Salt Lake City, San Diego, Boulder, Lincoln, Houston, Chicago, Urbana-Champaign, Ann Arbor, Pittsburgh, Ithaca, Atlanta, Washington D.C., Boston, Princeton.

4.2 Industry Data

The American Fact Finder website gave me a compact, state level summary of industry employment across all the states. This is for us to get a preliminary, broad industry overview. These table codes ECN_2017_US_00CCOMP1, ECN_2012_US_00CCOMP1, ECN_2007_US_00CCOMP1, and ECN_2002_US_00CCOMP1 record this data from 1997 to 2017.

The BLS has statistics ranging back to 1990 regarding employment levels in all metropolitan statistical areas within the United States (https://www.bls.gov/cew/downloadable-data-files.htm). This data records each county and MSA, and subdivides these regions' employment based on NAICS code. The NAICS code system allows us to track employment for specific industry types (e.g. the information industry), and splitting by MSA allows us to compare cities to one another.

4.3 Processing

There were significant logistical complications with processing the BLS data. The required information (segmentation by MSA, and segmentation by industry) was not recorded previous to 1990 – the data from 1975-1989 was effectively unusable because there was no way for me to map employment to industry, and also no categorization by MSA. In future work, more data prior to the NSFNET's construction would be valuable.

I process over 40GB of files data into a few MB of panel data, segmented by industry, containing the year, MSA/county (depending on granularity of analysis), and three outcomes of interest. I assign each MSA/county an appropriate "State" entry in anticipation of controlling for state fixed effects.

I also compose "centralization indices" for each of the outcomes of interest (more on that in the methodology section). This allows us to see not only absolute figures about the NAICS sector of interest, but how it performs relative to other industries in the MSA, and how that stacks up to the national economic picture.

I also add a column with "closest NSFNET node" and "distance to closest NSFNET node". I accomplish this by computing the center-of-mass centroids for each MSA/county polygon, computing the Haversine distance to each NSFNET node, and taking the minimum. Though ultimately an approximation, this facilitates future analysis of distance and its relationship to economic outcomes.

Specifics can be found in the appendix (github link included). It's complicated - significant Python and R skills are recommended.

5 Methodology

5.1 What industries do we care about?

We are particularly interested in how the information industry (NAICS code 51) evolved. This represents an interesting case study given that substantial portions of the industry simply did not exist prior to the 1990s (e.g. internet publishing), and other more traditional parts of the industry (e.g. broadcasting and media, or data processing services) were radically transformed with the introduction of the internet. We further segment the information industry into some components of interest; we segment to all industries in the second level of hierarchy in the NAICS classification system - industries 511 (publishing industries, non-internet), 512 (motion picture and sound), 515 (broadcasting, non-internet), 516 (internet publishing and broadcasting), 517 (telecommunications), and 518 (data processing and hosting). Don't ask me why the NAICS people skipped 513 and 514.

There are certainly other industries that were impacted, but to analyze them all is beyond the scope of this author.

5.2 Determining overall industry centralization/decentralization

We must first determine whether there has been centralization of the information industry or not at a national level. This allows us to do a 'pre-check' of the hypothesis to see if there is any kind of centralization occurring in the first place. It also directs the rest of the methodology – in a centralized world, we can measure centralization 'hotspots', but in a decentralizing world, we may need a different strategy to see where the industry is decentralizing towards. This is done through the use of the EG index [7, 9], as shown here:

$$\gamma_I^{EG} = \frac{\sum_{i=1}^{M}(s_i - x_i)^2 - (1 - \sum_{i=1}^{M} x_i^2)\sum_{j=1}^{N} z_j^2}{(1 - \sum_{i=1}^{M} x_i^2)(1 - \sum_{j=1}^{N} z_j^2)}$$

In this setup, N firms exist, and they choose among M locations. s_i is the firm's share of industry I employment in area i, x_i is the firm's share of total employment in area i, and the z_j are the sizes of the firms j of industry I.

Further presentation of results is in the results section; however, since a national-level centralization was found, we move towards future steps with the goal of looking towards centralization within certain MSAs.

5.3 Selecting and modifying our outcome variables

We are interested in two primary centralization concepts: first, the centralization of the information industry into specific MSAs, and second, the centralization of the MSAs' economy around the information industry. I ultimately propose that in addition to absolute figures (the number of establishments, the number of employees, and the average annual wage), we also construct of an index that encodes both centralization concepts.

On industry centralization - it is not obvious how to quantify centralization of an industry. Does it have to do with how many cities an industry operates in? Where they are hiring the most people? Whether the number of firms is lower than it used to be?

Consider this example: industry A used to have 1000 employees – 900 employees in one city, and 100 employees spread across 20 different cities. Now, it has changed its operations to have 250 employees across 4 different cities. Is the industry centralizing because it is operating in fewer cities, or decentralizing because there is a more even spread between the cities it operates in?

Unfortunately, there is no good "pre-made" answer. The EG index, as used in step one, uses share of total labor in an area as their determinant instead of output or firm numbers. This index has legitimacy; it appears often in the urban economics literature. But we can only use it for step 1, because it does not model firm choice. At a low level (say, the city level) the EG index for city X has no relation to the EG index for city Y. It treats industry in city X as an entirely different entity from industry in city Y, rather than treating the industry as a combined set of firms that can make the choice of what city to stay in. A high EG index for city X only shows that within the city, the industry has centralized (for example, moved all their operations to one building). There is unfortunately no commonly accepted indicator for this in the literature. It is actually an unsolved problem in urban economics, discussed in the Limitations section.

Furthermore, the EG index says nothing about an industry's relative standing to other industries, which is important to judge how an economy as a whole is centralizing around an industry. Let's say that after the 1994 treatment, there is a case of unambiguous industry centralization - 100 employees, split across 10 cities, eventually coalesce into 1000 employees in one city. Additionally, let's say that the city itself started out with 2000 workers from various industries, but as the information industry entered, many workers left, and now the information industry comprises 50% of the city's employees. Here, two kinds of centralization is happening - one is the industry's centralization into a city, and the other is the city's economy centralizing around the information industry.

I propose an index which encodes both of these centralization concepts. For a particular industry i from the set of industries I and any arbitrary outcome o, we use the 'location quotient' (LQ from here on out), defined as $\frac{LC}{NC}$ (local concentration / national concentration):

$$LQ_{industry,outcome,MSA} = \frac{LC_{industry,outcome,MSA}}{NC_{industry,outcome,MSA}}$$

$$LC_{i,o,m} = \frac{outcome\ o\ for\ industry\ i\ in\ MSA\ m}{\sum_{j \in I} outcome\ o\ for\ industry\ j\ in\ the\ MSA\ m}$$

$$NC_{i,o} = \frac{outcome\ o\ for\ industry\ i\ in\ the\ nation}{\sum_{j \in I} outcome\ o\ for\ industry\ j\ in\ the\ nation}$$

Here, the information industry outcomes are compared against industry outcomes in general within the MSA, allowing us to see if the MSA's economy centralizes around the information industry. Furthermore, this is compared to a national average of sorts, allowing us to see if a particular MSA has "more centralization" as compared to the rest of the nation.

As for the specific outcomes that we push through this "LQ" index, we pick the available measures of number of establishments, number of employees, and average annual wage.

Note that the LQ itself has an interpretation that is reliant on the "national average", but differences of LQs between MSAs (which our regressions will identify) does not depend on NC (since it is a constant independent of the specific MSA). Differencing two LQs just reflects differences in the industry's share of total firms between two MSAs.

5.4 Regression Strategy

Difference in differences is a potential strategy to consider in this case. There is a treatment that "activates" at a certain time, and we are interested in the before-after scenario.

The main problem, however, is the parallel trends assumption. In the absence of the treatment - the infusion of infrastructure and human capital - would these MSAs have developed like any other? I describe the issue, make some qualitative arguments in favor and against parallel trends, reference some quantitative tools that we can use to inspect parallel trends.

It is plausible - probable, even - that the assignment of treatments was nonrandom. The question is whether those unobservables are correlated to information industry growth. If they are, then there is no way to tell if the change in outcomes is a result of pre-existing qualities of that MSA, or a result of the treatment. If they are uncorrelated, it's fine - for instance, even if there was nonrandom assignment to MSAs that had nice Thai restaurants, it wouldn't really have an impact on long term information industry growth (well, hopefully not). We examine this question from a theoretical and empirical standpoint.

Evidence on knowledge spillovers from university to industry is mixed across different sectors;

evidence from Luc Anselin [1] suggests that spillovers only happen for certain industries, Audretch [2] suggest that the spillovers are not regionally defined, and Kantor and Whalley [11] identify spillovers as having an increasing impact over time. Most of the spillovers identified, however, have to do with private-public research networks which pay off over time. In our situation, many of the information industry sectors we are analyzing (publishing, broadcasting, etc) have little to do with research. There are industry sectors that could have benefited from research/human capital surrounding that research (telecommunications, data hosting/processing), but this advantage might not have "activated" without the treatment of regional network infrastructure to begin with.

Ultimately, we are exploring a historical context in which the internet exploded so rapidly that everyone was playing catch-up, businesses weren't really looking towards RD, and the human capital that was being selected on (supercomputing and research) don't relate in the short term to the commercial effects in broadcasting, publishing, etc. that followed. There is some logical basis to assume parallel trends, but also a logical basis to reject it. We turn to the data, and will present some graphs that explore the pre- and post- measurements of these outcome variables in treated MSAs relative to the conditions of the state that they are located in. We will also inspect coefficient values in a regression based event study (more details later in the methods section) to look for signifiers of parallel trends.

5.4.1 Basic pooled DID

We start with a basic DID model in which we pool together observations from the pre-treatment years and the post-treatment ones. This simple model looks like this:

$$y_{it} = \beta_0 + \beta_1 Post_t + \beta_2 Treated_i + \beta_{did}(Post_t * Connection_i) + \epsilon$$

where our parameter of interest is β_{did}, $Post_t$ and $Treated_i$ are dummies for being treated/being in a post-period, and y_{it} refers to any one of our six chosen outcome variables (avg employment, avg pay, num of establishments, and the location-quotient versions of all of those).

5.4.2 Adding state-time, and individual fixed effects

We use a two-way fixed effects model, using individual fixed effects and interacted state-time fixed effects rather than separate ones, to adjust for unobserved unit, group, and time-specific confounders in the same model.

There is much justification for adding state fixed effects. It is likely that states differ in their overall economic situation in a way that affects all the MSAs within the state; for instance, a state could have policies that make opening new businesses or changing business practices relatively easier. We don't necessarily want to compare an MSA in one state with an MSA in another.

Additionally, there is good reason to add time fixed effects. The problem with the pooled regression is that it treats all of the "post" years as if they were the same. This poses a few problems. First, we know that there are probably uniform economic shocks that affect many MSAs in certain years (e.g. The Great Recession). In those years, perhaps all economic measures are somewhat lower than those in previous years.

However, it is likely that the time-effect is not uniform by state. Moreover, the state-level heterogeneity discussed earlier is likely to vary over time. With this in mind, rather than have separate state and time fixed effects dummies, we interact the two. This allows for time effects to vary by state/state level effects to vary over time.

Finally, we include individual fixed effects. We assume that MSAs have different "initial conditions" that persist through time and create differences in outcome levels starting from the very beginning. Together with the interacted state-time fixed effect, this is the two-way fixed effects model. The regression specification looks like this:

$$y_{ist} = \beta_0 + \beta_{did}(Post_t * Treated_i) + \alpha_i + \gamma_{st} + \epsilon$$

Where α refers to the MSA fixed effect, γ refers to the interacted state-time effect, our parameter of interest is β_{did}, and y_{ist} refers to any one of our six chosen outcome variables.

5.5 Parallel trends analysis - event study regression

In addition to the graphs presented at the beginning of the section, we do an event study regression as follows:

$$y_{ist} = \beta_0 + \alpha_i + \gamma_{st} + \sum_{k=1990}^{k=2016} \beta_k(Treated_i * \mathbb{1}(t = k))$$

In which $(t = k)$ is an indicator for whether the observation is in year k and the other terms are the same as previously mentioned. We drop the dummy variable for 1995 due to collinearity, since 1995 was the year that the treatment (opening NSFNET to commercial use) was activated. This is just an event study with 28 time periods.

Each β_k is simply the average difference between treated and untreated groups at each time period 1990 through 2018. These β_k values cannot really tell us how significant the effect of the NSFNET treatment is overall – however, they can tell us how valid parallel trends are. If parallel trends hold, one would expect β_k values from 1990-1995 to be **relatively constant** - it is acceptable if they are nonzero, since that reflects an initial difference in levels, but we are hoping they remain mostly flat, to show that the treated MSAs are generally tracking together with other MSAs prior to the treatment. This should make sense if the NSFNET "treatment" had no bearing on industry/commercial operations.

However, after the treatment activates in 1995, we should observe markedly different beta values. Being able to show parallel trends in this way improves the validity of differences-in-differences.

5.6 Standard Error Clustering

R and STATA will both return incorrect p-values such as p=0.000 because by default, the SEs are clustered on each MSA observation at each time period, rather than by each MSA across all the relevant time periods. Make sure to cluster standard errors by MSA.

5.7 Extensions

5.7.1 Leveraging differences within the treated group

There are two hypothesized "competitive advantages" that the NSFNET opening conferred. The first has to do with infrastructure access, and the second has to do with human capital in the areas the NSFNET was being built in. We attempt to see which one is the primary effect by decomposing the treatment group into two separate treated groups: the "supercomputing centers" and the "educational centers". Both of these two treatment groups received funding to build out regional networks, but for very different reasons - the "supercomputing center" treated sites were selected on the basis of supercomputing research specifically (a hyper-specific human capital criteria that is not as likely to spill over into general commercial activity), whereas the "educational centers" treated sites were likely selected on broader criteria.

We re-run the regressions from above, but this time we restrict the treated group to only the "supercomputing centers" or "educational centers" and remove the observations corresponding to the other group. If we find that the supercomputing center sites did not benefit much from the treatment, we might be able to guess that the infrastructure itself might not have been much of a factor, and that it was more about the human capital or other institutional qualities.

5.7.2 Relationships with distance

We wonder if any of these measurements are correlated with distance from the central NSFNET node. We discuss why we care about distance, and what measures we should actually care about in this extension.

To bring back perspectives from urban economics, people tend to centralize around natural resources due to transport costs associated with moving those resources outwards. If we conceive of "connectivity" and "human capital regarding the internet" as a resource, then might we find spatial correlations that extend beyond the scale of an MSA? We construct MSA "centroids" that were provided with an NSFNET node, and we calculate each MSA's haversine distance to the nearest NSFNET "node".

As for what outcomes we actually want to correlate with distance, we are **not** particularly interested in absolute measures (pay, employment, etc). The same absolute figure for something like pay will mean something very different for two different areas that are both equidistant from the central node (e.g. if one area is relatively more developed than the other, or has a higher cost of living). We are more interested in the **relative** measures captured by the LQ; a different LQ-average pay or a different LQ-number of establishments tells us how dense/important the information industry is **relative to other industries** in that sector. That is the outcome of interest that actually defines something like industry centralization.

As for the actual regression, we consider a single "group" to be a treated MSA and all MSAs within a 100 kilometer radius. Then, with these distance measures, we measure the relationship between distance from a central node and the economic outcome. Each "group" probably experiences fixed effects; for instance, maybe one particular regional network is exceedingly good. Moreover, there are probably time-based fixed effects. Thus, we include state and time FE variables in a regression as follows. This is **not** a causal argument; I simply want to see if any correlations exist.

So, we run

$$y_{igt} = \beta_0 + \alpha_g + \gamma_t + \sum_{k=1990}^{k=2018} \beta_k Distance_i * \mathbb{1}(t = k) + \epsilon$$

where α_g captures the fixed effect for node-group g, γ_t captures the fixed effect for year t, and we retrieve one β_k for each year estimating the effect of distance on the outcome of interest. Once again, this is by **no means** causal. Repeat this for each industry outcome (LQ-pay, LQ-employment, LQ-number of establishments).

6 Results and Analysis

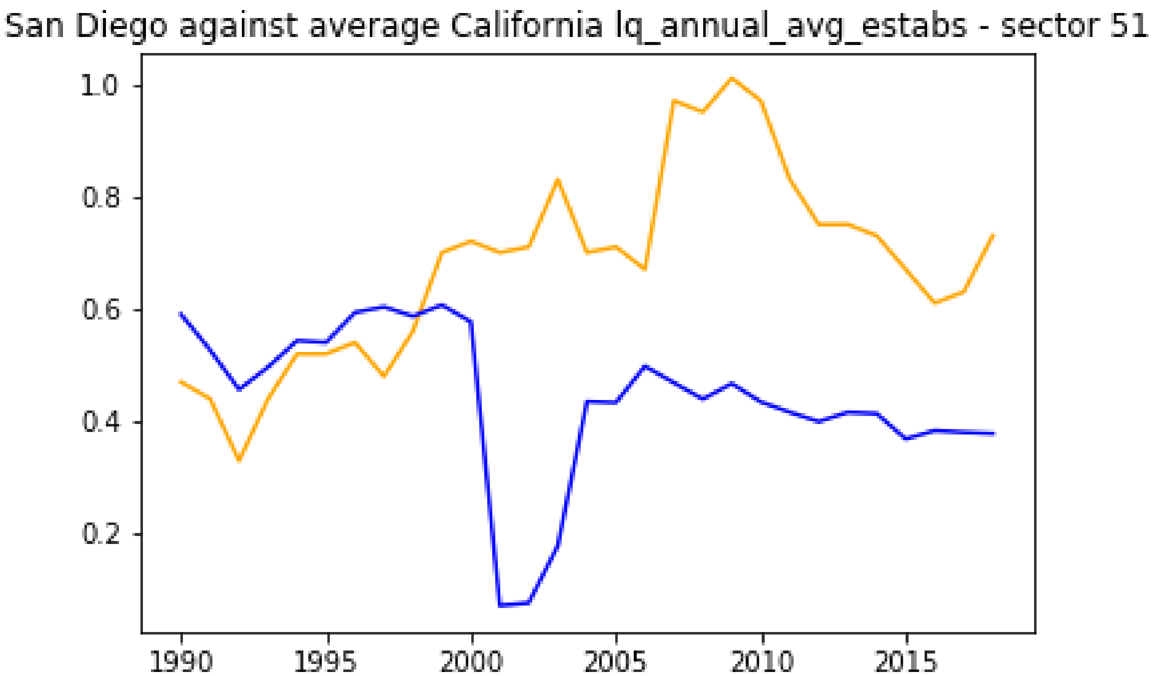
6.1 Overall centralization

We find that there is steady centralization occurring over time. The EG index was calculated to be 2.2342 in 1997, 2.6784 in 2002, 3.0235 in 2007, 3.1129 in 2012, and 3.3759 in 2017, showing a steady increase over time. The index on its own has no real-world interpretation, only the difference between the indices matters. Now we know that there is a definitive increase in centralization – now the question becomes where that centralization is taking place.

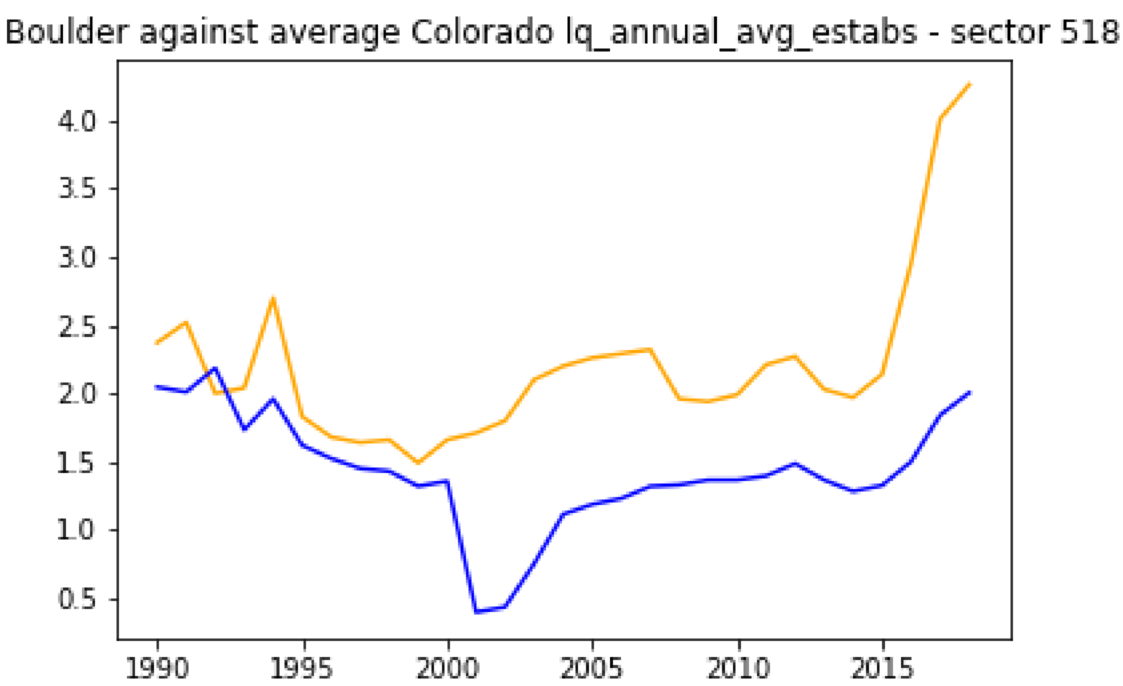
6.2 Initial visualizations for parallel trends

Here, I present some graphs which seem to support a parallel trends interpretation by presenting outcomes of an MSA in one industry against the state average outcome for that industry. Across 6 industry segments, 6 outcomes, and 13 MSAs of interest (which actually had data), there are 468 graphs total. Many of them appear to indicate parallel trends. All of them are linked in the appendix, but for the sake of brevity, I will post only a few as samples across different industry sectors, outcomes, and MSAs. Note: orange is the treated MSA, blue is the state average. Outcomes are annual_avg_estabs (number of establishments), annual_avg_emplvl (number of employees), avg_annual_pay (annual wages), all averaged across firms in the industry, and LQ (location quotient) versions. Some titles are cut off.

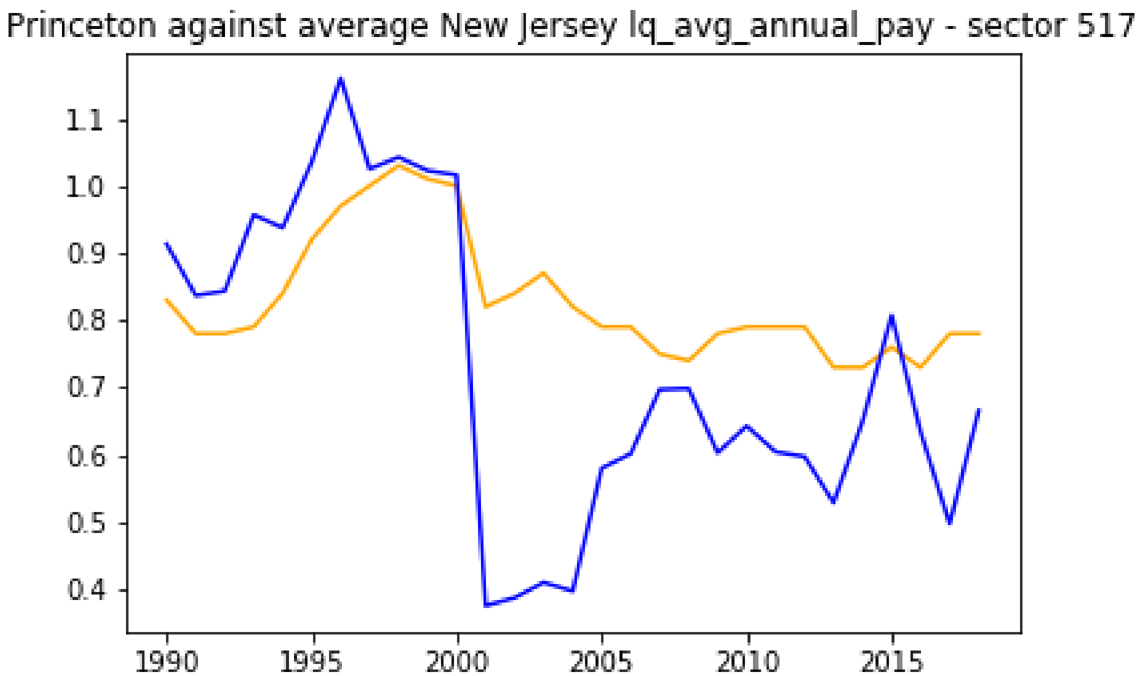
Industry 519 (News Syndicates, Libraries, Portals for Internet Publishing/Search):



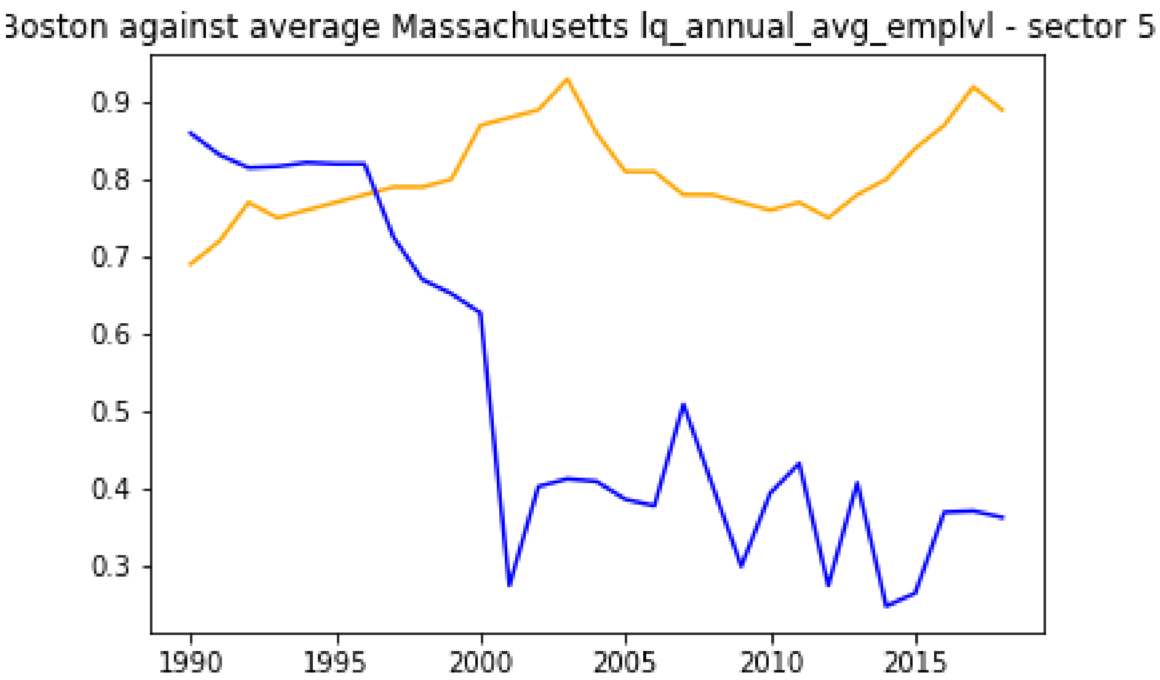
Industry 518 (Data hosting, processing, and related services):



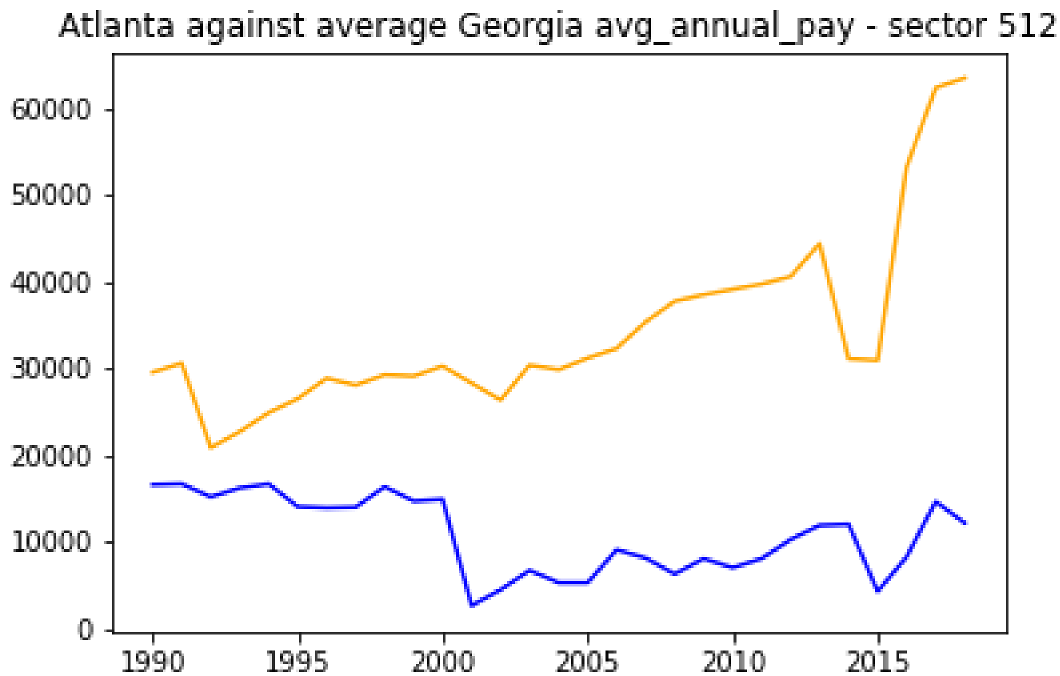
Industry 517 (Telecommunications):



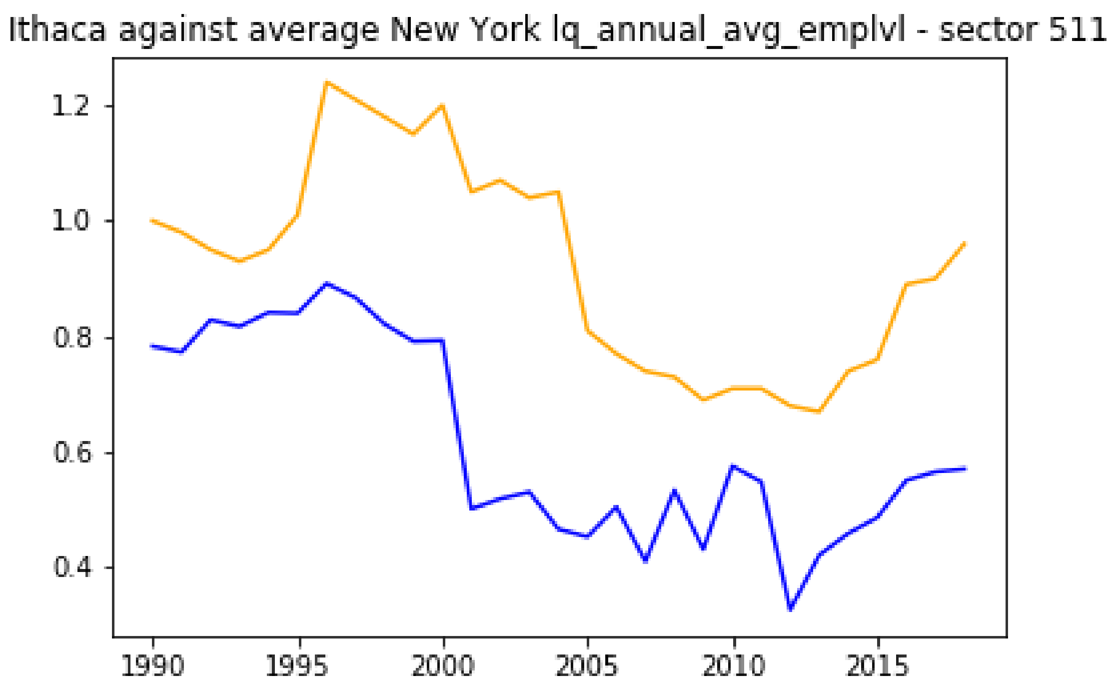
Industry 515 (Non-Internet Broadcasting):



Industry 512 (Motion Pictures and Sound):



Industry 511 (Non-Internet Publishing):



There is some indication that prior to the treatment (and for a year or two afterwards, which makes sense - the treatment may not have an immediate effect), the treated MSAs are functioning along similar lines. What's particularly interesting is that across multiple sectors and outcomes, there seems to be a massive drop in the outcome around 2001, and usually the state average does not recover while the treated MSA does.

This is a fascinating phenomenon for future research, and significantly affects the interpretation of our paper. It suggests that the NSFNET's infrastructure and human capital investment "inoculated" these regions in some sense against the economic shock of the dot com bubble bursting in 1999/2000. It also suggests that the relative "advantages" found in the treated MSAs with our regressions below aren't necessarily a result of faster growth than other regions, but rather, more sustainable growth that is resistant to shocks.

This idea is explored further later on; for now, we use the presentation of these parallel trends to move onto our DID regression results.

6.3 Regression results

6.3.1 Pooled two-period DID

There were 6 outcomes and 8 industries of interest for a total of 48 regressions, with the primary parameter of interest being the β_{did} coefficient. We present all the regressions in the appendix, separated by industry. The results are compressed into a brief table below, including only values for the β_{did} coefficient.

As a whole, it doesn't appear that effect sizes are very significant or consistent. Particularly of interest are the results for the absolute number of establishments in industries 516, 517, 518, and 519 (which is composed of news syndicates, libraries, and internet search and publishing portals); note that there are significant positive effects for the number of establishments in treated MSAs. (as an aside - for industry 519, the relevant driver of growth is probably not public libraries or news syndicates, and we can assume that internet publishing portals played a large part in that growth). This is particularly interesting in context of the results for industries 511, 512, and 515, which are **not** internet related; in treated MSAs, the effect sizes are both smaller, not consistent, and even negative.

It makes sense that standard errors on average employment and pay are large and that the effect sizes are inconsistent; there are so many other local variables that affect those things (e.g. broadly, the cost of living being high in a particular state or MSA, or general movement out of those MSAs). Given that this does not consider state-time fixed effects or unit-specific fixed effects, we approach these results with skepticism and move on to the fixed-effects version before considering an economic explanation.

Table 1: DID treatment parameter values for all industries and outcomes

	(avg empl)	(avg pay)	(num estabs)	(LQ-emp lvl)	(LQ-avg pay)	(LQ-num estabs)
51 (Information general)	-1092 (5218.6)	8285 (10101.5)	795.93(*) (326.67)	0.01656 (0.188)	−0.08263 (0.119)	−0.03980 (0.0552)
511 (publishing, non-internet)	-3069.3 (3280.6)	23298(*) (5703.9)	-58.05 (47.167)	0.4273 (0.29175)	0.06294 (0.06229)	-0.2291(*) (0.0714)
512 (motion picture, sound)	-109.2 (503.6)	3785 (2340.0)	46.73 (35.30)	0.04702 (0.06341)	0.03361 (0.03777)	0.0302 (0.06495)
515 (broadcasting, non-internet)	-7.753 (577.0)	12162 (6801.2)	1.280 (9.845)	0.3507(**) (0.09963)	-0.06517 (0.07829)	0.2502(**) (0.07075)
516 (internet publish/broadcasting)	198.44 (116.13)	10638 (8403)	25.1344(*) (7.5600)	0.8871 (0.5967)	0.17035 (0.12956)	0.82125 (0.4779)
517 (telecommunications)	-3026 (2449.0)	13216 (8051.5)	172.56(**) (50.714)	0.03115 (0.09368)	0.10453 (0.08629)	0.11735(.) (0.05851)
518 (data processing and hosting)	230.2 (1377.1)	30818.6(**) (7341)	89.032(*) (29.045)	0.2300 (0.18875)	0.24070(*) (0.08527)	-0.1222 (0.10249)
519 (other)	2428.8(*) (851.5)	28266(*) (8237)	136.95(**) (38.039)	1.7497(.) (0.7716)	-0.0538 (0.11315)	1.611(*) (0.04884)

Note: $^{\cdot}$p<0.1; *p<0.05; **p<0.01

6.3.2 State-time and unit-level fixed effects DID

Again, there were 6 outcomes and 8 industries of interest for a total of 48 regressions, with the main parameter of interest being β_{did}. All regression tables are in the appendix; a summarized version presenting the β_{did} values is presented at the end of this subsection.

Once again, most of the numbers are statistically insignificant. However, there are three columns that I'd like to focus on: average annual pay, the number of establishments, and the LQ-number of establishments.

The average pay seems to be higher for treated MSAs across all the information industry subsegments in a statistically significant way (except for motion picture/sound, sector 512). This makes some sense; 512 seems to be the sector that would be affected least by ICT technologies. However, some of the effect sizes seem too large to be reasonable; a $36000 difference in pay between a treated/nontreated MSA in sector 519 seems too good to be true.

The number of establishments and the LQ-number establishments tells a more interesting story about firm centralization. For the absolute number, when the fixed effects are included, the effects remain significant like in the pooled model, and most importantly, many of the relative levels remain unchanged - data processing/hosting, telecommunications, "other" (which, again, is likely to be driven by internet search and archiving) increase the most, while non-internet publishing, motion picture/sound, and non-internet broadcasting establishments tend to increase less.

However, the coefficients for the LQ-number of establishments require some more interpretation. We find that the publishing and data processing/hosting industries experience a significant decrease in LQ in the treated MSAs. What this means is that treated MSAs have a smaller proportion of firms in this sector than nontreated MSAs, which - contextualized with the absolute num estabs coefficient - means that for treated MSAs, their growth was outpaced by growth in other industries. The same is true for the data processing and hosting industry. In contrast, the internet publishing and broadcasting, telecommunications, and broadcasting industries experienced statistically significant absolute AND relative growth in treated MSAs more than in nontreated MSAs. This suggests that because of the treatment, these firms became relatively more dominant in these MSAs.

Results for non-internet broadcasting are also fairly interesting and unintuitive. One would expect that traditional broadcasting (TV, cable, etc) would be driven out by the rise of ICT. An alternative explanation is that in early years, traditional broadcasting was largely unthreatened by the nascent; though the internet could serve up static text and photo content that publishers excel in, it was unable to serve up longer form audio and video content in a convenient and accessible way (e.g. while driving). Thus, these businesses didn't really experience an initial competitive disadvantage.

Ultimately, from our fixed effects regression, we can really only say that pay went up across the board and that some industries appear to have become more central to their local economies and some didn't. Most of the effects are not significant.

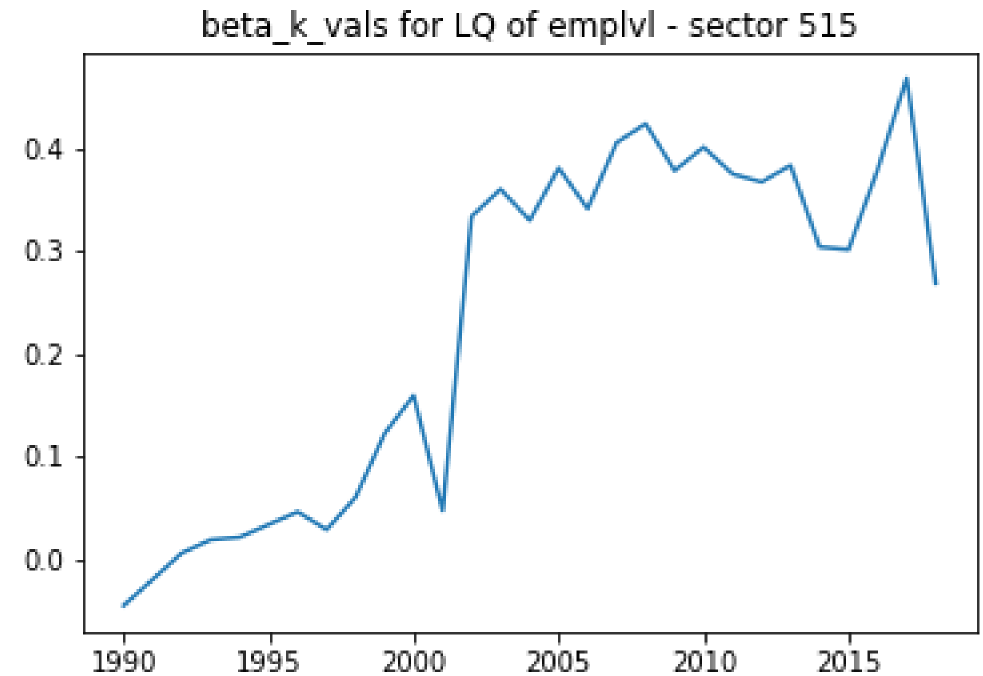
Table 2: Fixed Effect DID treatment parameter values for all industries and outcomes

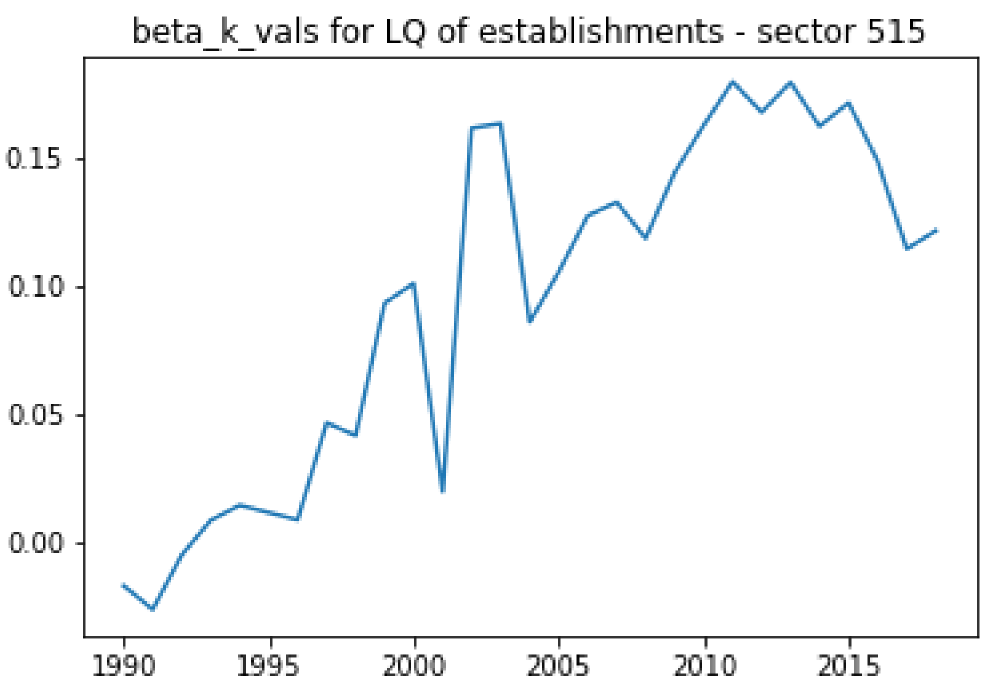
	(avg empl)	(avg pay)	(num estabs)	(LQ-emp lvl)	(LQ-avg pay)	(LQ-num estabs)
51 (Information general)	5690.2 (3680.2)	20986.9(**) (7409.2)	322.12(**) (121.13)	0.2837(*) (0.13125)	0.07664 (0.05050)	0.04073 (0.02636)
511 (publishing, non-internet)	-735.88 (2741.86)	24955.8(**) (3740.5)	27.362 (17.822)	0.41266 (0.26489)	0.05899(*) (0.02767)	-0.22644(**) (0.058271)
512 (motion picture, sound)	-25.815 (325.976)	911.86 (1768.57)	32.816(**) (10.466)	-0.013704 (0.059329)	-0.032598 (0.027117)	-0.0006651 (0.0454781)
515 (broadcasting, non-internet)	660.41(.) (390.55)	14867.3(**) (3975.3)	16.4310(**) (3.7207)	0.281437(**) (0.064666)	-0.036912 (0.029919)	0.120152(*) (0.048747)
516 (internet publish/broadcasting)	483.35(**) (151.31)	27663.4(**) (7666.1)	20.452 (13.905)	2.66766(**) (0.47255)	0.010655 (0.076866)	1.31041(**) (0.33978)
517 (telecommunications)	412.9 (3199.9)	14041.3(*) (7152.4)	194.848(**) (57.518)	0.096598 (0.081220)	0.096388 (0.082054)	0.119693(**) (0.037162)
518 (data processing and hosting)	-274.18 (1001.30)	18257.6(**) (3818.1)	56.314(*) (23.993)	-0.079395 (0.132287)	0.07016 (0.05106)	-0.194102(**) (0.064906)
519 (other)	2457.0(*) (1220.9)	36061(**) (11368)	81.971(.) (46.315)	0.57373 (0.82565)	0.0074317 (0.1012761)	0.059821 (0.145643)

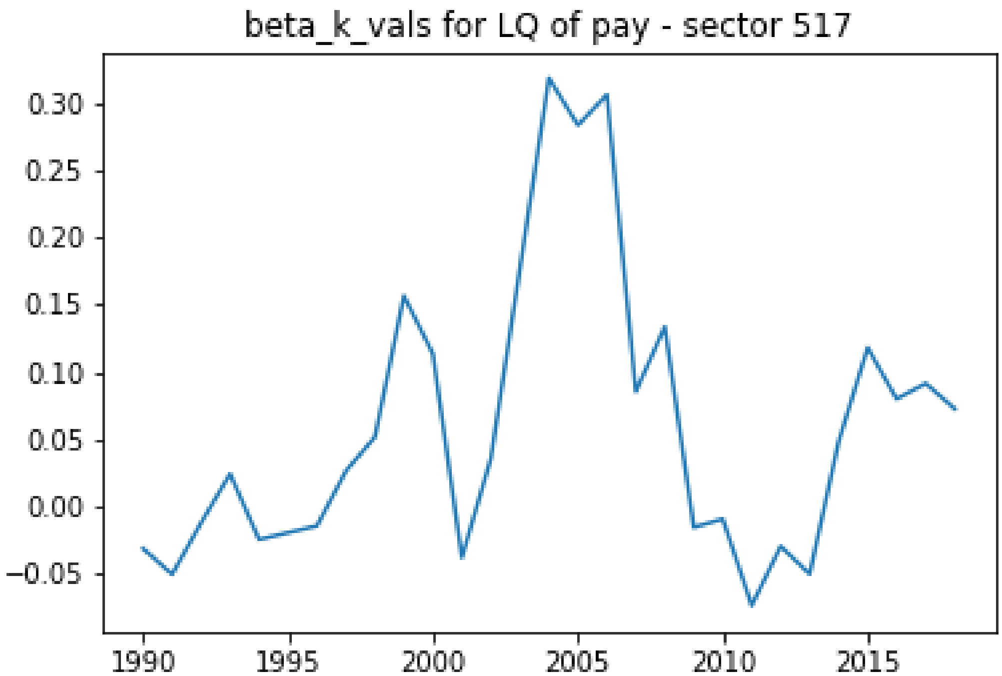
Note: (.)p<0.1; *p<0.05; **p<0.01

6.3.3 Event study regression

Between 8 industries and 6 different outcomes, there were 48 event studies to test the parallel trends assumption. We present only a few of the event studies (specifically for the LQ-outcomes) below in graph format by graphing the β_k values at each year for each industry sector and outcome; for 28 years, this is a total of 1344 β_k values to examine, so the regression tables are not included. For simplicity's sake, the graphs, which are much easier to read, are all linked digitally in the appendix, organized by industry and outcome.


beta_k_vals for LQ of emplvl - sector 515
0.0
0.1
0.2
0.3
0.4
1990
1995
2000
2005
2010
2015


beta_k_vals for LQ of establishments - sector 515
0.00
0.05
0.10
0.15
1990
1995
2000
2005
2010
2015



In a perfect parallel trends world, the β_k values pre-treatment would be constant. Here, we note that there is only weak support for parallel trends in some industry sectors; there is often a little bit of growth in the outcome variable prior to the treatment, but that growth accelerates drastically in the post-treatment phase. Thus, results in this paper should be interpreted with some skepticism.

6.3.4 Extension 1 - leveraging differences within the treatment group

The same regressions were run on different treatment subgroups (the MSAs that were chosen as supercomputing sites as well, and the sites that were chosen only for regional networks) in an attempt to decompose the effects. The regressions are presented in the appendix; since so many of the results are insignificant, there is very little we can say. The only interesting observation is that the DID coefficients across the industries tend to actually be pretty close for the supercomputing and regional network sites.

6.3.5 Extension 2 - Distance's effects on outcomes

We find statistically significant effects of essentially zero. Distance appears to have zero effect. This is itself an interesting result - it suggests that the information industry outcomes had nothing to do with the actual regional networks themselves; the "globalizing effect" of the new technology defied traditional methods of industry organization that organize themselves around a central resource. Regression tables for this are included in the appendix; they are largely uninteresting and filled with near-zero numbers. One of the three outcomes are presented below (with the fixed effects sizes removed for brevity); each of the "interact_XXXX" terms is the interaction specified between distance and year in section 5.7.2 of the paper:

```
Call:
lm(formula = lq_annual_avg_emplvl ~ factor(year) + factor(nearest_location) +
    interact_1990 + interact_1991 + interact_1992 + interact_1993 +
    interact_1994 + interact_1996 + interact_1997 + interact_1998 +
    interact_1999 + interact_2000 + interact_2001 + interact_2002 +
    interact_2003 + interact_2004 + interact_2005 + interact_2006 +
    interact_2007 + interact_2008 + interact_2009 + interact_2010 +
    interact_2011 + interact_2012 + interact_2013 + interact_2014 +
    interact_2015 + interact_2016 + interact_2017 + interact_2018,
    data = msa_df)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2281 -0.2441 -0.0320  0.1829  3.4911

Coefficients:
                                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                               7.691e-01  8.766e-02   8.774  < 2e-16 ***
interact_1990                            -3.151e-04  6.873e-04  -0.458 0.646705
interact_1991                            -2.974e-04  6.873e-04  -0.433 0.665266
interact_1992                            -7.080e-04  6.873e-04  -1.030 0.303010
interact_1993                            -1.068e-03  6.873e-04  -1.553 0.120465
interact_1994                            -1.284e-03  6.873e-04  -1.867 0.061907 .
interact_1996                            -1.344e-03  6.873e-04  -1.955 0.050589 .
interact_1997                            -1.529e-03  6.873e-04  -2.224 0.026202 *
interact_1998                            -1.300e-03  6.873e-04  -1.891 0.058636 .
interact_1999                            -1.501e-03  6.873e-04  -2.183 0.029083 *
interact_2000                            -1.765e-03  6.873e-04  -2.568 0.010250 *
interact_2001                            -1.392e-03  6.011e-04  -2.316 0.020624 *
interact_2002                            -5.135e-04  6.011e-04  -0.854 0.393034
interact_2003                            -7.914e-04  6.011e-04  -1.317 0.188037
interact_2004                            -4.582e-04  6.011e-04  -0.762 0.445976
interact_2005                            -5.846e-04  6.011e-04  -0.973 0.330853
interact_2006                             5.333e-05  6.011e-04   0.089 0.929311
interact_2007                            -2.865e-04  6.011e-04  -0.477 0.633618
interact_2008                            -1.077e-03  6.011e-04  -1.791 0.073316 .
interact_2009                            -6.621e-04  6.011e-04  -1.102 0.270737
interact_2010                            -1.452e-03  6.011e-04  -2.416 0.015740 *
interact_2011                            -8.057e-04  6.011e-04  -1.340 0.180181
interact_2012                            -1.097e-03  6.011e-04  -1.825 0.068107 .
interact_2013                            -1.882e-03  5.890e-04  -3.195 0.001408 **
interact_2014                            -2.475e-03  5.890e-04  -4.203 2.69e-05 ***
interact_2015                            -2.314e-03  5.890e-04  -3.929 8.65e-05 ***
interact_2016                            -1.817e-03  5.890e-04  -3.085 0.002049 **
interact_2017                            -3.159e-03  5.890e-04  -5.364 8.56e-08 ***
interact_2018                            -3.206e-03  5.890e-04  -5.443 5.51e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4067 on 4529 degrees of freedom
Multiple R-squared:  0.1947,    Adjusted R-squared:  0.1822
F-statistic: 15.64 on 70 and 4529 DF,  p-value: < 2.2e-16
```

The total lack of correlation between these information industry outcomes and distance means that the conventional urban economics "centralization around an urban center" narrative isn't a good explanation for the differences in number of firms and LQ-number of firms that we saw in the regressions. The conventional wisdom of human capital and resources localizing to minimize costs incurred by distance is likely not applicable to internet technologies in the first place, and the lack of correlation provides some evidence in favor of that interpretation.

7 Limitations, problems, future work

7.1 Data shortcomings

The data I have is limited because three sites did not report all the necessary data. Two NSFNET sites (Chicago, Urbana-Champaign) are in Illinois, and Illinois did not report all data to the BLS. Out of 26 years, there were only 5 years of available employment data. The same went for the site in College Park, Maryland. I tried looking for the information from other sources, but I could not find a good source that used the same reporting and categorization methods that the BLS used, so I had to leave these sites out of the data.

Additionally, for subsegments of the industries of interest, there was a lot of missing data. For instance, for industry code 516 (Internet Publishing), some MSAs simply did not report certain figures until the 2000s; for instance, Palo Alto would report annual average pay in this area, but indicate 0 establishments (misreported data, certainly). The missing data resists accurate imputation, because for many MSAs that aren't treated, there is no distinction made in the original files between whether something is 0 or missing, and no way to make the judgment call easily.

7.2 Lack of a good centralization index

The location quotient may not be a good indicator for industry centralization/agglomeration. This is perhaps the most difficult limitation to overcome, because there is no commonly accepted solution to this. As mentioned in the methodology section, it is exceedingly difficult to develop an index of how centralized an industry is. Scott Kominers at the Harvard Business School writes after examining 7 such indices in the literature, “It is possible that the optimal solution to the agglomeration index problem will be a combination of measurements... It is, unfortunately, not clear how to model such an index. We suspect that the answer may lie in a clever modeling application of a single statistical tool. We do not believe this tool has been found yet” [7]. The location quotient depends on what the MSA's employment distribution is, which may be a problem. If the MSA is suffering and hemorrhaging jobs, but the information industry is staying, then it appears that the industry is becoming more prominent in that area even though it is not really changing. This is a potential alternative explanation for results seen, which has a completely different economic story. This will be a limitation moving forward for future studies. Without a commonly accepted measure of centralization that can be compared across locations, it is difficult to examine this question.

7.3 Parallel trends - conceptual problems and the need for more covariates to condition on

Parallel trends faces a unique conceptual challenge in this specific economic context. Parallel trends is essentially a statement that we can use pre-treatment trends to guess what the counterfactual without treatment would be. The assumption is that if there are unobservables besides the treatment that would affect outcomes, they would already be affecting outcomes in the pre-treatment period and we would be able to see that through economic outcomes, despite being unable to observe those factors ourselves. The assumption is then made that in the counterfactual, non-treatment world, the unobservables continue to have the same level effect that they did in the pre-treatment period. The quantitative tests proposed earlier still assume this.

However, this assumption doesn't necessarily hold in this case. There is compelling reason to believe that existing unobservables wouldn't be "activated" until the post-treatment time period. In this scenario, the internet and internet-adjacent markets didn't truly experience their transformative growth until the post-treatment period. It could **appear** that unobservables from the pre-treatment period don't have an effect on growth, but that may only be because the mechanism for them to have an effect in the first place was blocked. Since that mechanism (the rise of the internet) would have happened independently of the treatment, it is difficult to say whether these pre-treatment trends can be projected forward as a counterfactual.

I propose a potential solution/extension - if we can gain access to covariates that measure things such as "human capital", "research activity", "institutional robustness", etc then we can condition our regressions on those covariates, thus adjusting for the previously **un**observables.

7.4 Dynamic treatment effects

We also have good reason to believe that the treatment effect we are testing for changes over time. The treatment (infrastructure+human capital investment) does not act instantly - it manifests in business practices and agglomeration effects that slowly pick up speed over time. We expect to see that the effect perhaps manifests in a small way in earlier years, but in a very large way in later years.

Another way to imagine this concept is to repeatedly run a two period DID - but instead of pooling the post-treatment observations together, restrict the post-treatment observations to a single year (starting at 1996). As a different DID model is run for each post-treatment year, the $\beta_{t,did}$ coefficient will probably get larger. However, this "hack" doesn't actually tell us anything about the treatment effect, and $\beta_{t,did}$ is likely to grow **anyways** (if both MSAs grow at a steady 5per year but have different starting points, the gap between them will grow larger and larger without bound, even though their "growth rate" is the same).

However, the basic idea still holds about dynamic treatment effects. To estimate this, we can turn to a DID extension proposed by Callaway and Sant'Anna [4] which they do a more complex version of the multiple two-period DID method proposed above, in which they estimate a treatment effect for individuals that have been treated for exactly e periods:

$$\tilde{\theta}_D(e) = \sum_{g=2}^{\mathcal{T}} \sum_{t=2}^{\mathcal{T}} \mathbf{1}\{t - g + 1 = e\} ATT(g,t) P(G = g | t - g + 1 = e),$$

and then average over all possible e to get an estimate of a "dynamic treatment effect" parameter:

$$\theta_D = \frac{1}{\mathcal{T} - 1} \sum_{e=1}^{\mathcal{T}-1} \tilde{\theta}_D(e).$$

(here, $ATT(g,t)$ is the average treatment effect for a particular time period t and a particular group g, which is computed using another method of their own in the paper). The math is complicated, but the work has been done and exists in an R package - in future work, this could be used to see how ICT's impact grows over time.

7.5 Future work

7.5.1 Explaining the lack of information industry outcome correlation with distance

As stated in the earlier section, the lack of correlation between distance from a nodal center and information industry outcomes means that the conventional model of urban centralization based on distance costs might not apply too well in this situation. Then, what does? Future work could conduct a more fine-grained analysis of business within the MSA. This might show an effect that our methods have hid; since our data is at the MSA level, it is certainly possible that this effect occurs to *some* degree, but just doesn't spill over very far. Future work could also work on new theoretical models for the new, "zero-transaction-cost" world brought forth by ICT.

7.5.2 Analyzing the dot-com bubble shock

Perhaps the most interesting takeaway from this whole paper wasn't the focus of it to begin with. Return back to the graphs presented in section 6.2 (I present a few more that were not shown previously).

All these graphs have a common trend - though the outcome variable for the state takes a massive plunge in 1999/2000, and seems to be permanently lowered by this plunge, the treated MSAs are robust to this shock, tending to be affected less and recover faster.

There are multiple potential reasons why. In the original hypothesis, I inferred that the treatment caused agglomeration effects that brought more information industry businesses to these hotspots. It could be that past a certain "critical mass" of firms and employees, an industry in a region becomes much more robust to shocks. Alternatively, since the treatment itself was focused on funding and supporting more robust institutions, it could be that institutions that are industry-adjacent can help support it during economic shocks. A final reason for this could be the **quality**, not merely the quantity, of firms; it could be that firms that are "first to the party" when it comes to new technology will be better managed, more experienced, and have stronger fundamentals rather than be simply following a technology trend.

This is a fascinating avenue for future research, and one that poses interesting questions for how to make industries more "crash resistant", especially since more of our economy will be built on new and emergent technologies in the future.

8 Conclusions

The results presented here are ultimately unclear and raise more questions than answers. From the regression results, it seems that the temporary human capital and infrastructure advantage conferred by the NSFNET's opening did have significant and persistent effects, specifically on the number of establishments in an area. In addition to the visual indications from the graphs presented in section 6.2, the regressions indicate that in treated MSAs, there generally tends to be more information firms both in absolute terms and relative to firms in other industries. While positive effects are identified for factors like pay and employment, those effects tended not to be particularly significant, and don't vary too consistently across industry subgroups in a way that invites a logical explanation. Moreover, the lack of correlation of these outcomes with distance prevents us from applying basic urban economics theories such as centralization-due-to-transction-costs to this domain. Finally, multiple big assumptions made about parallel trends throw these results into question. Even if the graphs and the event study regressions were completely reliable, there is still the major conceptual assumption discussed in the limitations section.

In the process of answering the primary question, this paper ran into many more. More consistent historical data is needed. More work is needed to develop an authoritative measure of centralization that can model firm choice. More novel DID methods can be applied to model dynamic treatment effects. More work needs to be done on the part of the author to figure out how to do spatial econometrics. More work needs to be done in urban economics to see how ICT's effects concur with or flaunt dominant theories of urban development.

The internet truly did take the world by storm; perhaps no good/service has propagated so quickly and so deeply into our lives the history of the world. It makes sense that we are still coming to terms with its effects on our economies and the mechanisms by which it operates. Far more work is needed to understand ICT's past effects, so we can understand its future potential to restructure and reshape our economies.

9 Works Cited

References

[1] Anselin, Luc and Varga, Attila. *Geographical Spillovers and University Research: A Spatial Econometric Perspective.* https://doi.org/10.1111/0017-4815.00142 Growth and Change, A Journal of Urban and Regional Policy, Dec 2002.

[2] Audretch, David and Feldman, Maryann. *Knowledge Spillovers and the Geography of Innovation.* Handbook of Urban and Regional Economics, Vol. 4, May 2003.

[3] Bade, F.J. *Urban specialization in the internet age – empirical findings for Germany.* Kiel Institute for the World Economy, 2004.

[4] Callaway, Brantly and SantAnna, Pedro. *Difference-in-Differences with Multiple Time Periods.* https://dx.doi.org/10.2139/ssrn.3148250. Mar 2019

[5] Computer History Museum. *Internet History of the 1980s.* https://www.computerhistory.org/internethistory/1980s/. 2020.

[6] Cybertelecom - Federal Internet Law and History: NSFNET Acceptable Use Policy. *Internet History :: NSFNET.* http://www.cybertelecom.org/notes/nsfnet.htmaup. Oct 2019.

[7] Duke Kominers, Scott. *Measuring Agglomeration.* 2008.

[8] Eichengreen, Barry; Lafarguette, Romain; and Mehl, Arnaud. *Cables, Sharks and Servers: Technology and the Geography of the Foreign Exchange Market.* http://dx.doi.org/10.3386/w21884, National Bureau for Economic Research, Jan 2016.

[9] Ellison, Glen and Glaeser, Edward L. *Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach.* Journal of Political Economy, 1997.

[10] Gaspar, Jess and Glaeser, Edward L. *Information Technology and the Future of Cities.* Journal of Urban Economics, 1998.

[11] Kantor, Shawn and Whalley, Alexander. *Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks.* https://doi.org/10.1162/REST_a_00357. The Review of Economics and Statistics, Mar 2014.

[12] Leamer, Edward and Storper, Michael. *The economic geography of the internet age.* National Bureau for Economic Research, Aug 2001.

[13] Merit Networks. *NSFNET: A Partnership for High-Speed Networking, Final Report* https://web.archive.org/web/20170202190223/https://www.merit.edu/wiki/NSFNET_final.pdf. 1995.

[14] National Research Council. *Funding a Revolution: Government Support for Computing Research* The National Academies Press, 1999.

[15] National Science Foundation. *Internet: Changing The Way We Communicate.* https://web.archive.org/web/20081209095319/http://www.nsf.gov/about/history/nsf0050/pdf/internet.pdf.

[16] O'Sullivan, Arthur. *Urban Economics* McGraw-Hill Education, 2006.

[17] Sinai, Todd. *Geography and the Internet: Is the Internet a substitute for cities?* Journal of Urban Economics, July 2004.

[18] Townsend, Anthony. *The Internet and the rise of the new network cities.* Environment and Planning: Planning and Design, Feb 1999.

10 Appendix

10.1 Data Preprocessing and Replication Code

See github.com/gaoag/senior-honors-thesis/ and read the README.md file for an explanation of the different notebooks, scripts, processed data, and how to replicate the results. R and Python skills recommended.

10.2 Parallel Trends Visualizations

See github.com/gaoag/senior-honors-thesis/ for the folder containing all the parallel trends visualizations. With 13 MSAs, 6 industries of interest, and 6 outcomes per industry for a total of 468 graphs, there are too many to include in the appendix at once.

10.3 Pooled DID Regression Results

The regression summaries are posted below. They are also available in text file format at github.com/gaoag/senior-honors-thesis/

10.3.1 Industry 51 (Information, General)

```
[[1]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper     DF
(Intercept)            0.83106   0.013710 60.6178 1.545e-147  0.80405  0.85807 240.10
treated_general_dummy  0.36171   0.081560  4.4349  9.652e-04  0.18255  0.54087  11.18
post_dummy            -0.04799   0.009691 -4.9522  1.343e-06 -0.06708 -0.02891 252.82
did                   -0.03980   0.055205 -0.7209  4.846e-01 -0.15990  0.08030  12.17

Multiple R-squared:  0.05348 ,  Adjusted R-squared:  0.05323
F-statistic: 16.55 on 3 and 387 DF,  p-value: 3.861e-10

[[2]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error   t value   Pr(>|t|) CI Lower CI Upper     DF
(Intercept)            0.79647    0.02072  38.44684 1.373e-104  0.75566   0.8373 240.10
treated_general_dummy  0.30046    0.09884   3.03990  1.106e-02  0.08334   0.5176  11.18
post_dummy            -0.17089    0.01645 -10.38674  2.807e-21 -0.20329  -0.1385 252.82
did                    0.01656    0.18800   0.08809  9.312e-01 -0.39244   0.4256  12.17

Multiple R-squared:  0.04632 ,  Adjusted R-squared:  0.04606
F-statistic: 42.36 on 3 and 387 DF,  p-value: < 2.2e-16

[[3]]

Call:
estimatr::lm_robust(formula = lq_avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error  t value   Pr(>|t|)  CI Lower CI Upper     DF
(Intercept)            0.86317   0.007503 115.0365 5.468e-212  0.848387   0.8779 240.10
treated_general_dummy  0.09468   0.041264   2.2944  4.209e-02  0.004033   0.1853  11.18
post_dummy            -0.20945   0.011153 -18.7792  7.425e-50 -0.231418  -0.1875 252.82
did                   -0.08263   0.119866  -0.6894  5.035e-01 -0.343406   0.1781  12.17

Multiple R-squared:  0.06568 ,  Adjusted R-squared:  0.06543
F-statistic:   124 on 3 and 387 DF,  p-value: < 2.2e-16
```

```
[[4]]

Call:
estimatr::lm_robust(formula = annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)             238.56      84.08   2.837 4.939e-03    72.93   404.18 240.10
treated_general_dummy   625.61     264.24   2.368 3.697e-02    45.16  1206.06  11.18
post_dummy               71.03      12.98   5.474 1.060e-07    45.48    96.59 252.82
did                     795.93     326.67   2.436 3.111e-02    85.25  1506.61  12.17

Multiple R-squared:  0.07241 ,  Adjusted R-squared:  0.07216
F-statistic:  15.5 on 3 and 387 DF,  p-value: 1.534e-09

[[5]]

Call:
estimatr::lm_robust(formula = annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper     DF
(Intercept)               8016     3006.6  2.6660 0.008197   2093.0  13938.4 240.10
treated_general_dummy    20634     9191.3  2.2450 0.045931    443.7  40824.8  11.18
post_dummy               -1212      725.1 -1.6714 0.095879  -2640.1    216.1 252.82
did                      -1092     5218.6 -0.2092 0.837777 -12444.8  10261.6  12.17

Multiple R-squared:  0.02232 ,  Adjusted R-squared:  0.02206
F-statistic: 4.061 on 3 and 387 DF,  p-value: 0.00735

[[6]]

Call:
estimatr::lm_robust(formula = avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper     DF
(Intercept)              26686      375.5 71.0748 2.951e-163    25946    27426 240.10
treated_general_dummy    10604     2162.1  4.9047  4.466e-04     5855    15354  11.18
post_dummy               11801      717.0 16.4588  7.253e-42    10389    13213 252.82
did                       8285    10101.5  0.8202  4.279e-01   -13691    30261  12.17

Multiple R-squared:  0.0662 ,   Adjusted R-squared:  0.06595
F-statistic: 101.5 on 3 and 387 DF,  p-value: < 2.2e-16
```

10.3.2 Industry 511 (Publishing, non-Internet)

```
[[1]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)            0.82049    0.02348  34.938 5.829e-87  0.77418  0.86680 199.467
treated_general_dummy  0.90251    0.18714   4.823 1.034e-03  0.47704  1.32798   8.714
post_dummy            -0.06723    0.01591  -4.225 3.531e-05 -0.09859 -0.03586 214.481
did                   -0.22911    0.07148  -3.205 1.058e-02 -0.39051 -0.06770   9.104

Multiple R-squared:  0.1202 ,   Adjusted R-squared:   0.12
F-statistic: 15.89 on 3 and 387 DF,  p-value: 9.213e-10

[[2]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)             0.8879    0.04241  20.938 3.029e-52   0.8043   0.9715 199.467
treated_general_dummy   0.5645    0.12375   4.562 1.481e-03   0.2832   0.8459   8.714
post_dummy             -0.3053    0.03394  -8.997 1.256e-16  -0.3722  -0.2384 214.481
did                     0.4273    0.29175   1.465 1.767e-01  -0.2315   1.0861   9.104

Multiple R-squared:  0.1142 ,   Adjusted R-squared:  0.1139
F-statistic: 45.07 on 3 and 387 DF,  p-value: < 2.2e-16

[[3]]

Call:
estimatr::lm_robust(formula = lq_avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper      DF
(Intercept)            0.81325    0.01127  72.130 7.449e-145  0.79102   0.8355 199.467
treated_general_dummy  0.23419    0.05861   3.995  3.342e-03  0.10093   0.3675   8.714
post_dummy            -0.31066    0.01409 -22.050  4.775e-57 -0.33843  -0.2829 214.481
did                    0.06294    0.06229   1.010  3.384e-01 -0.07772   0.2036   9.104

Multiple R-squared:  0.1221 ,   Adjusted R-squared:  0.1218
F-statistic: 177.4 on 3 and 387 DF,  p-value: < 2.2e-16
```

```
[[4]]

Call:
estimatr::lm_robust(formula = annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)              53.24      6.819   7.807 3.235e-13    39.79    66.69 199.467
treated_general_dummy   577.69    142.599   4.051 3.079e-03   253.49   901.89   8.714
post_dummy               17.80     14.612   1.218 2.244e-01   -11.00    46.61 214.481
did                     -58.05     47.167  -1.231 2.492e-01  -164.57    48.46   9.104

Multiple R-squared:  0.2279 ,   Adjusted R-squared:  0.2277
F-statistic: 6.016 on 3 and 387 DF,  p-value: 0.0005163

[[5]]

Call:
estimatr::lm_robust(formula = annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)             1736.7      234.5  7.4060 3.588e-12   1274.3  2199.14 199.467
treated_general_dummy  18239.0     5021.8  3.6320 5.771e-03   6821.9 29656.19   8.714
post_dummy              -410.4      218.1 -1.8821 6.118e-02   -840.2    19.41 214.481
did                    -3069.3     3280.6 -0.9356 3.736e-01 -10477.6  4339.00   9.104

Multiple R-squared:  0.2938 ,   Adjusted R-squared:  0.2936
F-statistic: 6.838 on 3 and 387 DF,  p-value: 0.0001686

[[6]]

Call:
estimatr::lm_robust(formula = avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper      DF
(Intercept)              23023      388.5  59.261 1.368e-128    22257    23789 199.467
treated_general_dummy    15440     2643.3   5.841  2.790e-04     9431    21450   8.714
post_dummy                8404      849.2   9.895  2.942e-19     6730    10077 214.481
did                      23298     5703.9   4.085  2.674e-03    10418    36179   9.104

Multiple R-squared:  0.1154 ,   Adjusted R-squared:  0.1151
F-statistic: 48.87 on 3 and 387 DF,  p-value: < 2.2e-16
```

10.3.3 Industry 512 (Motion and Audio)

```
[[1]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)             0.6943    0.03263 21.2779 6.328e-43   0.6297   0.7589 122.28
treated_general_dummy   0.3173    0.07796  4.0692 1.751e-03   0.1462   0.4883  11.31
post_dummy             -0.1635    0.02121 -7.7096 2.509e-12  -0.2055  -0.1216 134.39
did                     0.0302    0.06495  0.4649 6.507e-01  -0.1121   0.1725  11.41

Multiple R-squared:  0.06811 ,  Adjusted R-squared:  0.06782
F-statistic: 37.79 on 3 and 386 DF,  p-value: < 2.2e-16

[[2]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)            0.56513    0.03074 18.3846 5.333e-37  0.50428   0.6260 122.28
treated_general_dummy  0.12287    0.06081  2.0207 6.762e-02 -0.01051   0.2563  11.31
post_dummy            -0.25579    0.02618 -9.7708 2.295e-17 -0.30757  -0.2040 134.39
did                    0.04702    0.06341  0.7416 4.733e-01 -0.09192   0.1860  11.41

Multiple R-squared:  0.04453 ,  Adjusted R-squared:  0.04423
F-statistic: 52.15 on 3 and 386 DF,  p-value: < 2.2e-16

[[3]]

Call:
estimatr::lm_robust(formula = lq_avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error  t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)            0.46760    0.01848  25.2966 2.020e-50  0.43101   0.5042 122.28
treated_general_dummy  0.09110    0.03572   2.5502 2.647e-02  0.01274   0.1695  11.31
post_dummy            -0.24358    0.01723 -14.1351 2.304e-28 -0.27767  -0.2095 134.39
did                    0.03361    0.03777   0.8898 3.919e-01 -0.04916   0.1164  11.41

Multiple R-squared:  0.09395 ,  Adjusted R-squared:  0.09367
F-statistic: 105.2 on 3 and 386 DF,  p-value: < 2.2e-16
```

```
[[4]]

Call:
estimatr::lm_robust(formula = annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper     DF
(Intercept)             107.30      55.86   1.921  0.05706    -3.27   217.88 122.28
treated_general_dummy   206.03     111.26   1.852  0.09031   -38.02   450.08  11.31
post_dummy              -39.90      29.23  -1.365  0.17452   -97.71    17.91 134.39
did                      46.73      35.30   1.324  0.21147   -30.62   124.07  11.41

Multiple R-squared:  0.02756 ,  Adjusted R-squared:  0.02726
F-statistic: 3.504 on 3 and 386 DF,  p-value: 0.01555

[[5]]

Call:
estimatr::lm_robust(formula = annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper     DF
(Intercept)             1235.2      654.2  1.8882  0.06137   -59.75   2530.1 122.28
treated_general_dummy   2089.0     1196.1  1.7465  0.10778  -534.68   4712.6  11.31
post_dummy              -525.5      400.3 -1.3127  0.19153 -1317.24    266.3 134.39
did                     -109.2      503.6 -0.2169  0.83215 -1212.80    994.4  11.41

Multiple R-squared:  0.01127 ,  Adjusted R-squared:  0.01096
F-statistic:  3.42 on 3 and 386 DF,  p-value: 0.01741

[[6]]

Call:
estimatr::lm_robust(formula = avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)              13922      754.4  18.454 3.794e-37    12429    15416 122.28
treated_general_dummy     6166     1634.9   3.772 2.945e-03     2580     9753  11.31
post_dummy               -3053      638.1  -4.784 4.456e-06    -4315    -1791 134.39
did                       3785     2340.0   1.617 1.331e-01    -1343     8912  11.41

Multiple R-squared:  0.03195 ,  Adjusted R-squared:  0.03165
F-statistic:  16.8 on 3 and 386 DF,  p-value: 2.799e-10
```

10.3.4 Industry 515 (Broadcasting, non-Internet)

```
[[1]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)             1.3604    0.04842  28.095 1.170e-66  1.26483  1.45597 174.492
treated_general_dummy  -0.6737    0.10664  -6.317 3.863e-04 -0.92546 -0.42184   7.050
post_dummy             -0.1427    0.03101  -4.600 7.711e-06 -0.20383 -0.08149 189.076
did                     0.2502    0.07075   3.536 8.447e-03  0.08526  0.41514   7.531

Multiple R-squared:  0.02712 ,  Adjusted R-squared:  0.02683
F-statistic: 15.82 on 3 and 386 DF,  p-value: 1.009e-09

[[2]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)             1.1835    0.03559  33.258 1.945e-77   1.1133  1.25377 174.492
treated_general_dummy  -0.3264    0.12638  -2.583 3.611e-02  -0.6248 -0.02799   7.050
post_dummy             -0.4575    0.03122 -14.653 5.561e-33  -0.5191 -0.39591 189.076
did                     0.3507    0.09963   3.520 8.639e-03   0.1185  0.58299   7.531

Multiple R-squared:  0.04968 ,  Adjusted R-squared:  0.0494
F-statistic:    72 on 3 and 386 DF,  p-value: < 2.2e-16

[[3]]

Call:
estimatr::lm_robust(formula = lq_avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error  t value   Pr(>|t|) CI Lower CI Upper      DF
(Intercept)            0.82905    0.01342  61.7809 1.557e-120   0.8026   0.8555 174.492
treated_general_dummy  0.31132    0.07859   3.9612  5.375e-03   0.1257   0.4969   7.050
post_dummy            -0.29579    0.01512 -19.5582  2.678e-47  -0.3256  -0.2660 189.076
did                   -0.06517    0.07829  -0.8324  4.308e-01  -0.2477   0.1173   7.531

Multiple R-squared:  0.08839 ,  Adjusted R-squared:  0.08813
F-statistic: 142.4 on 3 and 386 DF,  p-value: < 2.2e-16
```

```
[[4]]

Call:
estimatr::lm_robust(formula = annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)             24.281      3.939   6.164 4.781e-09   16.506   32.056 174.492
treated_general_dummy   75.569     22.748   3.322 1.260e-02   21.855  129.282   7.050
post_dummy              -5.027      1.069  -4.700 4.990e-06   -7.136   -2.917 189.076
did                      1.280      9.845   0.130 8.999e-01  -21.671   24.231   7.531

Multiple R-squared:  0.1515 ,   Adjusted R-squared:  0.1513
F-statistic: 12.78 on 3 and 386 DF,  p-value: 5.612e-08

[[5]]

Call:
estimatr::lm_robust(formula = annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error  t value Pr(>|t|) CI Lower CI Upper      DF
(Intercept)           1082.254      384.5  2.81457 0.005446    323.3  1841.16 174.492
treated_general_dummy 3166.571      913.1  3.46809 0.010316   1010.6  5322.51   7.050
post_dummy            -320.867      113.7 -2.82251 0.005275   -545.1   -96.62 189.076
did                     -7.753      577.0 -0.01344 0.989629  -1353.0  1337.45   7.531

Multiple R-squared:  0.0479 ,   Adjusted R-squared:  0.04762
F-statistic: 14.49 on 3 and 386 DF,  p-value: 5.835e-09

[[6]]

Call:
estimatr::lm_robust(formula = avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)              24171      654.0  36.958 1.926e-84    22880    25461 174.492
treated_general_dummy    18099     4352.0   4.159 4.183e-03     7823    28375   7.050
post_dummy                5199      827.2   6.285 2.207e-09     3568     6831 189.076
did                      12162     6801.2   1.788 1.139e-01    -3694    28017   7.531

Multiple R-squared:  0.07555 ,  Adjusted R-squared:  0.07528
F-statistic: 19.64 on 3 and 386 DF,  p-value: 7.171e-12
```

10.3.5 Industry 516 (Internet Publishing)

```
[[4]]

Call:
estimatr::lm_robust(formula = annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)             7.3137     0.5555 13.1660 5.466e-08    6.088    8.539 10.794
treated_general_dummy   7.7363     5.3655  1.4419 2.498e-01   -9.893   25.365  2.841
post_dummy              0.5927     1.9612  0.3022 7.677e-01   -3.683    4.868 11.940
did                    25.1344     7.5600  3.3247 3.886e-02    2.296   47.973  3.308

Multiple R-squared:  0.09876 ,  Adjusted R-squared:  0.09737
F-statistic: 4.209 on 3 and 280 DF,  p-value: 0.006211

[[5]]

Call:
estimatr::lm_robust(formula = annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper     DF
(Intercept)             176.25      57.13  3.0851  0.01059    50.22   302.27 10.794
treated_general_dummy    31.35     115.21  0.2722  0.80407  -347.17   409.88  2.841
post_dummy             -105.01      59.83 -1.7552  0.10482  -235.44    25.42 11.940
did                     198.44     116.13  1.7087  0.17744  -152.39   549.26  3.308

Multiple R-squared:  0.03334 ,  Adjusted R-squared:  0.03186
F-statistic: 2.078 on 3 and 280 DF,  p-value: 0.1033

[[6]]

Call:
estimatr::lm_robust(formula = avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)              28042       2603 10.7729 4.134e-07    22299    33784 10.794
treated_general_dummy     4314       4789  0.9007 4.375e-01   -11421    20048  2.841
post_dummy              -12509       2417 -5.1745 2.352e-04   -17779    -7239 11.940
did                      10638       8403  1.2660 2.873e-01   -14746    36022  3.308

Multiple R-squared:  0.03458 ,  Adjusted R-squared:  0.03309
F-statistic: 12.63 on 3 and 280 DF,  p-value: 9.069e-08
```

```
[[1]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper     DF
(Intercept)            1.87382     0.4467  4.1944 0.001562   0.8883   2.8594 10.794
treated_general_dummy -0.07782     0.5414 -0.1437 0.895287  -1.8567   1.7010  2.841
post_dummy            -1.18701     0.4188 -2.8345 0.015115  -2.0999  -0.2741 11.940
did                    0.82125     0.4779  1.7184 0.175620  -0.6225   2.2650  3.308

Multiple R-squared:  0.09474 ,  Adjusted R-squared:  0.09335
F-statistic: 7.589 on 3 and 280 DF,  p-value: 6.75e-05

[[2]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper     DF
(Intercept)             1.3163     0.4545  2.8963  0.01480   0.3137  2.31890 10.794
treated_general_dummy  -0.4228     0.5687 -0.7435  0.51389  -2.2912  1.44565  2.841
post_dummy             -0.9895     0.4351 -2.2741  0.04223  -1.9381 -0.04093 11.940
did                     0.8871     0.5967  1.4867  0.22554  -0.9155  2.68960  3.308

Multiple R-squared:  0.04793 ,  Adjusted R-squared:  0.04646
F-statistic: 3.224 on 3 and 280 DF,  p-value: 0.02307

[[3]]

Call:
estimatr::lm_robust(formula = lq_avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)            0.89961    0.06150  14.628 1.855e-08   0.7639   1.0353 10.794
treated_general_dummy -0.01361    0.08557  -0.159 8.843e-01  -0.2948   0.2675  2.841
post_dummy            -0.66081    0.05636 -11.724 6.585e-08  -0.7837  -0.5379 11.940
did                    0.17035    0.12956   1.315 2.723e-01  -0.2211   0.5618  3.308

Multiple R-squared:  0.1435 ,   Adjusted R-squared:  0.1422
F-statistic: 77.73 on 3 and 280 DF,  p-value: < 2.2e-16
```

10.3.6 Industry 517 (Telecommunications)

```
[[1]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)            1.07148    0.02729  39.265 4.395e-96  1.01767  1.12529 200.498
treated_general_dummy -0.09448    0.08695  -1.087 3.064e-01 -0.29221  0.10325   8.698
post_dummy            -0.12419    0.02310  -5.376 1.973e-07 -0.16972 -0.07866 215.272
did                    0.11735    0.05851   2.006 7.556e-02 -0.01482  0.24952   9.088

Multiple R-squared:  0.01101 ,  Adjusted R-squared:  0.01074
F-statistic: 9.728 on 3 and 387 DF,  p-value: 3.343e-06

[[2]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)            0.82578    0.02938 28.1109 1.746e-71   0.7679   0.8837 200.498
treated_general_dummy  0.25911    0.18682  1.3870 2.000e-01  -0.1657   0.6840   8.698
post_dummy            -0.21264    0.02628 -8.0928 4.248e-14  -0.2644  -0.1609 215.272
did                    0.03115    0.09368  0.3325 7.471e-01  -0.1805   0.2427   9.088

Multiple R-squared:  0.02758 ,  Adjusted R-squared:  0.0273
F-statistic: 23.69 on 3 and 387 DF,  p-value: 4.269e-14

[[3]]

Call:
estimatr::lm_robust(formula = lq_avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value   Pr(>|t|) CI Lower  CI Upper      DF
(Intercept)            1.01572   0.008843 114.856 6.671e-185   0.9983  1.033157 200.498
treated_general_dummy -0.05394   0.022062  -2.445  3.794e-02  -0.1041 -0.003769   8.698
post_dummy            -0.28798   0.016628 -17.319  1.137e-42  -0.3208 -0.255203 215.272
did                    0.10453   0.086299   1.211  2.563e-01  -0.0904  0.299467   9.088

Multiple R-squared:  0.05152 ,  Adjusted R-squared:  0.05126
F-statistic:   102 on 3 and 387 DF,  p-value: < 2.2e-16
```

```
[[4]]

Call:
estimatr::lm_robust(formula = annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)              54.40     12.908   4.215 3.778e-05    28.95    79.86 200.498
treated_general_dummy   243.58     75.689   3.218 1.099e-02    71.45   415.71   8.698
post_dummy               25.34      4.813   5.265 3.386e-07    15.85    34.83 215.272
did                     172.56     50.714   3.403 7.731e-03    58.01   287.11   9.088

Multiple R-squared:  0.155 ,    Adjusted R-squared:  0.1548
F-statistic: 14.72 on 3 and 387 DF,  p-value: 4.267e-09

[[5]]

Call:
estimatr::lm_robust(formula = annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper      DF
(Intercept)               3158     1132.2   2.789 0.005798      925   5390.0 200.498
treated_general_dummy    15248     5116.4   2.980 0.016032     3612  26883.2   8.698
post_dummy               -1500      909.4  -1.650 0.100487    -3293    292.3 215.272
did                      -3026     2449.0  -1.235 0.247651    -8557   2506.3   9.088

Multiple R-squared:  0.1459 ,   Adjusted R-squared:  0.1457
F-statistic:  4.97 on 3 and 387 DF,  p-value: 0.002143

[[6]]

Call:
estimatr::lm_robust(formula = avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper      DF
(Intercept)              36009      332.9 108.183 8.983e-180    35353    36666 200.498
treated_general_dummy     7769     1369.3   5.674  3.450e-04     4655    10883   8.698
post_dummy                5155      986.0   5.228  4.042e-07     3212     7099 215.272
did                      13216     8051.5   1.641  1.348e-01    -4971    31403   9.088

Multiple R-squared:  0.03628 ,  Adjusted R-squared:  0.036
F-statistic: 19.32 on 3 and 387 DF,  p-value: 1.075e-11
```

10.3.7 Industry 518 (Data Hosting and Processing)

```
[[1]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)             0.9856    0.03309  29.786 1.289e-52   0.9200   1.0512 103.635
treated_general_dummy   0.7690    0.20624   3.729 4.364e-03   0.3055   1.2326   9.397
post_dummy             -0.2992    0.02951 -10.140 1.353e-17  -0.3577  -0.2408 113.607
did                    -0.1222    0.10249  -1.192 2.612e-01  -0.3512   0.1068   9.783

Multiple R-squared:  0.1505 ,   Adjusted R-squared:  0.1502
F-statistic: 43.05 on 3 and 383 DF,  p-value: < 2.2e-16

[[2]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)             0.9414    0.12581   7.483 2.466e-11  0.69191   1.1909 103.635
treated_general_dummy   0.5950    0.24099   2.469 3.459e-02  0.05329   1.1366   9.397
post_dummy             -0.5254    0.09415  -5.581 1.648e-07 -0.71193  -0.3389 113.607
did                     0.2300    0.18875   1.219 2.516e-01 -0.19182   0.6518   9.783

Multiple R-squared:  0.07651 ,  Adjusted R-squared:  0.0762
F-statistic: 22.53 on 3 and 383 DF,  p-value: 1.865e-13

[[3]]

Call:
estimatr::lm_robust(formula = lq_avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)            0.87085    0.01693  51.438 1.338e-75  0.83727   0.9044 103.635
treated_general_dummy  0.09339    0.04710   1.983 7.735e-02 -0.01247   0.1992   9.397
post_dummy            -0.45996    0.02191 -20.989 6.738e-41 -0.50337  -0.4165 113.607
did                    0.24070    0.08527   2.823 1.844e-02  0.05015   0.4313   9.783

Multiple R-squared:  0.1013 ,   Adjusted R-squared:  0.101
F-statistic: 162.9 on 3 and 383 DF,  p-value: < 2.2e-16
```

```
[[4]]

Call:
estimatr::lm_robust(formula = annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper      DF
(Intercept)             50.922     20.493   2.485  0.01456   10.281    91.56 103.635
treated_general_dummy  159.761     70.924   2.253  0.04960    0.347   319.17   9.397
post_dummy              -7.594      9.552  -0.795  0.42826  -26.518    11.33 113.607
did                     89.032     29.045   3.065  0.01223   24.122   153.94   9.783

Multiple R-squared:  0.1269 ,   Adjusted R-squared:  0.1266
F-statistic:  3.67 on 3 and 383 DF,  p-value: 0.01245

[[5]]

Call:
estimatr::lm_robust(formula = annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper      DF
(Intercept)             1703.6      563.4  3.0239 0.003146    586.4   2820.9 103.635
treated_general_dummy   5191.3     2394.3  2.1682 0.057036   -190.2  10572.9   9.397
post_dummy              -966.7      423.6 -2.2822 0.024337  -1805.8   -127.6 113.607
did                      230.2     1377.1  0.1672 0.870649  -2847.3   3307.7   9.783

Multiple R-squared:  0.1584 ,   Adjusted R-squared:  0.1581
F-statistic: 4.429 on 3 and 383 DF,  p-value: 0.004468

[[6]]

Call:
estimatr::lm_robust(formula = avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper      DF
(Intercept)            27820.5        684 40.6724 1.576e-65    26464    29177 103.635
treated_general_dummy   8739.9       1931  4.5263 1.285e-03     4400    13080   9.397
post_dummy              -326.9       1288 -0.2537 8.002e-01    -2879     2226 113.607
did                    30818.6       7341  4.1981 1.924e-03    14412    47225   9.783

Multiple R-squared:  0.07198 ,   Adjusted R-squared:  0.07167
F-statistic: 10.48 on 3 and 383 DF,  p-value: 1.214e-06
```

10.3.8 Industry 519 (Internet Search and Publishing Portals, News Syndicates, Libraries, Other)

```
[[1]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)              2.543     0.4086   6.224 8.029e-07   1.7081   3.3784 29.498
treated_general_dummy   -1.072     0.6188  -1.732 1.273e-01  -2.5371   0.3940  6.944
post_dummy              -1.703     0.3812  -4.469 9.509e-05  -2.4804  -0.9265 31.478
did                      1.611     0.4884   3.299 1.218e-02   0.4684   2.7544  7.378

Multiple R-squared:  0.02861 ,  Adjusted R-squared:  0.0282
F-statistic: 11.55 on 3 and 376 DF,  p-value: 2.921e-07

[[2]]

Call:
estimatr::lm_robust(formula = lq_annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)             2.0155     0.5062  3.9817 0.0004108  0.98095   3.0500 29.498
treated_general_dummy  -0.5727     0.7071 -0.8099 0.4448841 -2.24742   1.1021  6.944
post_dummy             -1.6481     0.4530 -3.6383 0.0009717 -2.57139  -0.7248 31.478
did                     1.7497     0.7716  2.2675 0.0558020 -0.05617   3.5555  7.378

Multiple R-squared:  0.06135 ,  Adjusted R-squared:  0.06095
F-statistic: 8.429 on 3 and 376 DF,  p-value: 1.959e-05

[[3]]

Call:
estimatr::lm_robust(formula = lq_avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error  t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)            0.76504    0.04803  15.9293 4.975e-16   0.6669   0.8632 29.498
treated_general_dummy  0.50496    0.15054   3.3544 1.232e-02   0.1484   0.8615  6.944
post_dummy            -0.55534    0.04547 -12.2140 1.757e-13  -0.6480  -0.4627 31.478
did                   -0.05387    0.11315  -0.4761 6.478e-01  -0.3187   0.2109  7.378

Multiple R-squared:  0.2035 ,   Adjusted R-squared:  0.2031
F-statistic:  68.6 on 3 and 376 DF,  p-value: < 2.2e-16
```

```
[[4]]

Call:
estimatr::lm_robust(formula = annual_avg_estabs ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper     DF
(Intercept)             37.236     17.628  2.1124 0.043232     1.21    73.26 29.498
treated_general_dummy   18.779     25.647  0.7322 0.488010   -41.97    79.52  6.944
post_dummy              -6.243      8.735 -0.7147 0.480054   -24.05    11.56 31.478
did                    136.950     38.039  3.6002 0.007995    47.93   225.97  7.378

Multiple R-squared:  0.08558 ,  Adjusted R-squared:  0.08519
F-statistic: 4.994 on 3 and 376 DF,  p-value: 0.002082

[[5]]

Call:
estimatr::lm_robust(formula = annual_avg_emplvl ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error  t value Pr(>|t|) CI Lower CI Upper     DF
(Intercept)              894.8      643.2  1.39125  0.17455   -419.7   2209.3 29.498
treated_general_dummy     20.8      721.1  0.02885  0.97779  -1687.0   1728.6  6.944
post_dummy              -511.5      385.6 -1.32666  0.19417  -1297.4    274.4 31.478
did                     2428.8      851.5  2.85239  0.02327    436.0   4421.6  7.378

Multiple R-squared:  0.03653 ,  Adjusted R-squared:  0.03612
F-statistic: 4.998 on 3 and 376 DF,  p-value: 0.002069

[[6]]

Call:
estimatr::lm_robust(formula = avg_annual_pay ~ treated_general_dummy +
    post_dummy + did, data = msa_df, clusters = area_fips)

Standard error type:  CR2

Coefficients:
                      Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper     DF
(Intercept)              19176       1307  14.671 4.327e-15    16505    21848 29.498
treated_general_dummy    17812       4953   3.596 8.902e-03     6081    29543  6.944
post_dummy               -1467       1297  -1.130 2.669e-01    -4111     1178 31.478
did                      28266       8237   3.432 1.011e-02     8989    47544  7.378

Multiple R-squared:  0.1193 ,   Adjusted R-squared:  0.1189
F-statistic: 7.688 on 3 and 376 DF,  p-value: 5.346e-05
```

10.4 State-Time Fixed Effects Regression Results

There aren't really any other parameters of interest here besides the β_{did} variable; the size of the fixed effects or the intercept don't matter much. All the coefficients are provided in the original section; replication code is in the github repository.

10.5 Event Study Graphs

See github.com/gaoag/senior-honors-thesis/event-study-graphs/ for the folder containing all the event study graphs and the RData file containing all the coefficients (it is nested as a hash table that you can access by loading the RData file and selecting a particular industry code (e.g. loaded_file[['511']]).

There are too many coefficients (8 industries, 28 years, 6 outcomes) to present here efficiently.

10.6 Extension 1 Regression Results (separating supercomputing treated sites from educational treated sites)

See github.com/gaoag/senior-honors-thesis/

10.7 Extension 2 Regression Results (correlating outcomes with distance)

See github.com/gaoag/senior-honors-thesis/