CESifo WORKING PAPERS

9090
2021

May 2021

The Impact of Online Competition on Local Newspapers: Evidence from the Introduction of Craigslist

Milena Djourelova, Ruben Durante, Gregory J. Martin

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9090*

The Impact of Online Competition on Local Newspapers: Evidence from the Introduction of Craigslist

Abstract

How does competition from online platforms affect the organization, performance, and editorial choices of newspapers? And what are the implications of these changes for the information voters are exposed to and for political accountability? We study these questions using the staggered introduction of Craigslist - the world's largest online platform for classified advertising - across US counties between 1995 and 2009. This setting allows us to separate the effect of competition for classified advertising from other changes brought about by the Internet, and to compare newspapers that relied more or less heavily on classified ads ex ante. We find that, following the entry of Craigslist, local papers experienced a significant decline in the number of newsroom and management staff. Cuts in editorial staff disproportionately affected reporters covering politics. These organizational changes led to a reduction in news coverage of politics and political corruption, and resulted in a decline in newspaper readership which was not compensated by increased news consumption on other media. Finally, we find some evidence that reduced news coverage of politics was associated with lower voter turnout, and more party-line voting for both citizens and politicians.

JEL-Codes: L820, L860, D720.

Keywords: newspapers, internet, advertising, political accountability.

Milena Djourelova
UPF, BGSE, IPEG
Barcelona / Spain
milena.djourelova@upf.edu

Ruben Durante
ICREA-UPF, BGSE, IPEG, CEPR
Barcelona / Spain
ruben.durante@upf.edu

Gregory J. Martin
Stanford GSB / CA / USA
gjmartin@stanford.edu

May 2021

We thank Tim Besley and seminar participants at UPF, Bristol, Bocconi, LSE, the Harris School of Public Policy and the European Commission (ECFIN), as well as participants in the Workshop on Elections, Democracy, and Populism at the University of Bologna, the CEPR Virtual IO seminar, SAEe and the Monash-Warwick-Zurich Text-as-Data Workshop for helpful comments. David Ampudia Vicente, Marco Lo Faso, and Elliot Motte provided excellent research assistance. Ruben Durante acknowledges financial support from the European Union's Horizon 2020 research and innovation program [Grant 759885].

1. Introduction

The Internet has profoundly changed the environment in which newspapers operate. In particular, due to increased competition from online platforms, newspapers' advertising revenues have sharply declined, forcing many outlets to downsize and drastically rethink their business model and organization (McChesney and Nichols, 2011). These changes are likely to impact the quality of news reporting and the provision of political information (Starkman, 2014; Peterson, 2020), and may have important implications for political accountability given the key role played by local newspapers in informing citizens about the conduct of their representatives (Gentzkow et al., 2011; Snyder and Strömberg, 2010).

Despite the potentially grave consequences of these transformations for the future of journalism, rigorous evidence on the impact of online competition on newspapers' organization and editorial choices is surprisingly scant. This is primarily due to the difficulty of separating the effect of online competition from other technological and socioeconomic changes brought about by the Internet, which may affect both the demand and the supply side of the newspaper market in other ways.

To overcome this limitation, in this paper we investigate the effect on newspapers of the introduction of *Craigslist* (henceforth CL), the world's largest online platform for classified ads. CL disrupted the market for classified ads, a formerly lucrative niche for newspapers (Seamans and Zhu, 2014; Kroft and Pope, 2014).[1] Tracking the expansion of CL across the U.S. between 1995 and 2009, we examine how the entry of a local CL website affected the organization and editorial decisions of local newspapers, and ultimately, the behavior of local voters and politicians.

The expansion of CL across the U.S. provides an ideal setting for several reasons. First, CL's slow and staggered expansion over a period of 15 years, combined with the limited geographic scope of each local CL website, generates significant variation over time and across space in the degree of online competition for classified ads faced by local newspapers. Second, since CL websites do not feature news content or display advertising, the entry of a CL website represents a specific shock to revenues from classified ads but leaves other market conditions unaffected. In addition, CL's narrow focus on classified ads provides an important source of heterogeneity, since the entry of CL should disproportionately affect newspapers that relied more heavily on classified ads *ex ante*. Finally, with a few exceptions in the biggest cities, ads on CL are free of charge, and most local websites do not generate profit for the

[1] According to the Newspaper Association of America, as of 2000 classified advertising accounted for 40% of US newspapers' total advertising revenues. In that year advertising revenues amounted to US$49 billion, compared to US$11 billion of circulation revenues.

company. The lack of a clear profit maximization motive at the company[2] reduces concerns that the timing of CL's entry might have been driven by strategic considerations related to the conditions of local newspaper or classified ad markets.[3]

We first validate our measure of CL roll-out and track the dynamics of take-up at the local level using data on the browsing behavior of a panel of online consumers. We find that visits to the Craigslist domain increase by about 20% after the opening of a local website, and that, consistent with network effects, this increase is gradual over several years.

To estimate the effects of CL's entry at the newspaper level, we compare the evolution of the outcomes of interests before and after the entry of a local CL website between areas with and without access to a local CL website, controlling for the quality of local broadband Internet and for a range of demographic and socio-economic conditions. In addition, we exploit variation across newspapers, including those headquartered in the same state or media market, in the reliance on classified ads prior to the entry of CL. Our proxy for the degree of reliance is the presence of a dedicated classified ads manager in the newspaper's staff.[4]

Using this approach, and exploiting comprehensive data on the organization of over 1,500 newspapers between 1995 and 2010, we first show that, following the entry of CL in a given market, the number of active local newspapers does not decline. However, the entry of CL is associated with a significant decrease in the number of staff of local newspapers. An event-study analysis shows the absence of any pre-existing trends, and indicates that the effect of CL on staffing levels takes a few years to fully materialize. In line with our hypothesis, the effect is mainly driven by newspapers that, prior to the entry of CL, relied more heavily on classified ads. Staff cuts affected both managerial and editorial positions. However, cuts in editorial staff disproportionately affected reporters covering politics, while other areas such as sports and entertainment remained largely unaffected.

We then test to what extent such organizational changes affected newspaper content, with particular regard to news coverage of politics. First, applying keyword searches to the entire corpus of articles published in over 800 newspapers, we compute the number of mentions of local House representatives, and document that, following the entry of CL in a given area, news coverage of local representatives in affected papers decreased. We then estimate a topic model on a random sample of two million articles and find evidence of a decline in the relevance of topics related to politics, with other topics such as sport, entertainment,

[2] CL founder Craig Newmark was sued by eBay, which held a stake in CL, in 2010 for failing in his fiduciary duty to maximize shareholder returns.

[3] Indeed, we document that the timing of the entry is not correlated with the characteristics of local newspapers once population and the quality of local Internet connection are controlled for.

[4] To corroborate the validity of this proxy, we document that the presence of classified ads managers is indeed strongly correlated with the share of pages devoted to classified ads prior to CL entry (see Appendix Table B1).

and crime remaining unaffected. Finally, we study how the entry of CL affects the ability of local papers to inform readers about the (mis)conduct of politicians. Specifically, we construct a measure of news coverage of public corruption measured by the frequency of mentions of corruption charges and phrases associated with investigative journalism (from Hamilton, 2016). We find that following the entry of CL news coverage of corruption in local papers declined significantly, even after controlling for actual corruption, proxied by the number of federal prosecutions for public corruption recorded in the state.

Next we examine how readers responded to changes in content. First, we document that, in the years after the entry of CL, local newspapers experienced a sharp decline in circulation. We further explore this aspect using data from two large-scale surveys on media consumption. The results confirm that, following the entry of CL, local respondents are less likely to report reading a (non-national) newspaper. Interestingly, this trend is primarily driven by individuals that are relatively *less* likely to be interested in classifieds, and therefore cannot be merely due to lower demand for print ads. Crucially, the decline in newspaper readership is not compensated by increased news consumption through other sources (i.e., online, radio, TV), and is hence likely to result in individuals being generally less informed.

Finally, we study how reduced news coverage and awareness of politics affect the behavior of local voters and politicians. Indeed, previous work indicates that more informed voters participate more and are better able to monitor and hold politicians accountable (Snyder and Strömberg, 2010). Regarding voters, we find that in areas affected by the entry of CL, turnout in presidential elections tend to decrease by about 1 percentage point. In addition, voters in these areas are significantly less likely to vote for different parties in concurrent legislative and presidential elections, consistent with a greater tendency to rely on national partisan cues when less information about local politicians is available (Moskowitz, 2021). For politicians, we find that members of Congress from areas most affected by CL - i.e., from districts that coincide more with newspaper markets where CL website entered (Snyder and Strömberg, 2010) - become more aligned to their party leadership in roll-call votes. This evidence supports the role of newspapers in maintaining the "electoral connection" between politicians and their district constituencies (Mayhew, 1974), and indicate that when changes in the newspaper market weaken this link, politicians are more likely to conform to party pressures.

Taken together our results indicate that the impoverishment of local newspapers due to competition from online platforms can jeopardize their ability to inform citizens about politics, with potentially detrimental effect for political participation, ideological polarization, and political accountability. This evidence supports the concerns expressed by some regulators that newspapers' financial distress, due to lower advertising revenues, may threaten quality

reporting and pluralism (FCC, 2016).

Our paper contributes to several streams of literature. First, it relates to previous work on the impact of the Internet on political participation, electoral outcomes, and public policies (Falck et al., 2014; Campante et al., 2017; Gavazza et al., 2019; Larcinese and Miner, 2018; Manacorda and Tesei, 2020). Exploiting arguably exogenous variation in access to broadband or wireless technology, these studies gauge the combined effect of the various changes brought about by the Internet on both the demand and supply side of the political market. In this respect, the novelty of our paper is to isolate the effect of the Internet - and specifically of digital advertising platforms - on legacy media and its implications for content quality, readership, and political outcomes.

Second, our paper relates to the literature on the link between media, information, voters' participation, and political accountability. This includes the seminal studies by Besley and Burgess (2002) and Snyder and Strömberg (2010) which document that media coverage of local politicians make voters' more informed and participative and politicians more responsive to the demands of their constituents. Especially related to our work is a recent study by Gao et al. (2020) which documents how the closure of local newspapers affected local public finances, leading to an increase in borrowing costs, government wages and deficits, and discretionary spending. The authors interpret these findings as evidence of the importance of local newspapers in monitoring local government. Our paper contributes to this literature by showing how the impoverishment of local newspapers and the resulting changes in organisation, content, and readership, can affect the "virtuous cycle" of accountability, making voters less participative and more ideologically polarised, and politicians less responsive to their constituents.

Third, our paper relates to previous work on the effect of online platforms on traditional media's business strategies in two-sided markets also exploiting the entry of CL. The most notable example is the work by Seamans and Zhu (2014) which shows that, after the entry of CL in an area, local newspapers experienced lower classified-ad rates and circulation, and higher display-ad rates and subscription prices. Our paper builds and expands upon these findings by documenting the entire "chain of events" triggered by the entry of CL, and its profound implications for newspaper staff, editorial priorities, news content, and, ultimately, political outcomes.[5]

Finally, our results relate to previous evidence on the impact of the introduction of new media and advertising technologies on traditional media outlets. For example, Angelucci

[5] Other work exploits the experiment created by CL's expansion across the US or particular design features to study questions related to matching efficiency in labor and housing markets (Kroft and Pope, 2014), as well the impact of online personal/erotic ads on sexually transmitted diseases and violence against women (Cunningham et al., 2019; Chan and Ghose, 2014).

et al. (2020) look at the effect of the introduction of TV on newspapers in the US, and find that, in areas where TV was rolled out earlier on, newspapers experienced lower circulation and advertising revenues, and published fewer pages and less original content. In a longer-term historical perspective, Hamilton (2004) and Petrova (2011) argue that the growth of the print advertising market in the late 19th century was essential to the emergence of an independent (non-partisan) press. Our paper studies the flip side of this question: what happens to newspapers' political coverage when advertising profits are competed away by new technology.

More closely related to our work is a recent paper by Bhuller et al. (2020) on how the adoption of broadband Internet affected traditional print media in Norway. Their findings indicate that Internet adoption accounts for about 40% of the overall reduction in circulation experienced by Norwegian newspapers between 2000 and 2010. This effect applied to national newspapers but not to local ones which managed to cut costs and to avoid losses also by changing their content from tabloid to more serious news. Our empirical setting is different in that it allows us to separate the effect of a specific shock to the advertising market from the many other changes brought about by the Internet. Furthermore, in addition to market outcomes, we also study how this shock affects news coverage of both national and local politics, and, ultimately the behavior of voters and politicians.

The remainder of the paper is organized as follows. Section 2 provides some background information about Craigslist and its expansion. Section 3 describes the data used in the analysis, while section 4 discusses the empirical strategy. Section 5 presents the results . Section 6 concludes.

2. Background

Craiglist.org (CL) is the world's largest online platform for classified ads. It was founded in San Francisco in 1995, and served only the Bay Area until 2000, when it began to gradually expand to other U.S. locations. CL initially opened new local websites in big cities such as Boston, New York, and Chicago. Over time, it expanded to smaller markets covering 115 locations in 2005, 331 in 2008, and 416 today.

Consistently ranked among the top 20 U.S. websites by traffic,[6] the CL platform has a simple and intuitive layout which has remained largely unchanged over time. Ads on CL websites are organized in a number of sections, the most relevant ones being housing, jobs, items for sale, professional services, and personal. Ads can include pictures of the item as well as the exact location of the seller who can be contacted by potential buyers directly through the platform. CL websites only host classified ads and do not include any display ads or news

[6] https://www.similarweb.com/top-websites/united-states/

content. Ads on CL websites are generally posted free of charge, with a few exceptions for brokered apartment rentals in New York and job posts for employers in some major cities.[7] This business model reflects the unconventional views of CL's founder Craig Newmark who prioritised providing a useful service to local communities over profit maximization, and always opposed listing the company on the stock market.[8]

Being cheaper and more efficient than traditional newspaper ads, CL ads became very popular among users, rapidly disrupting the lucrative market for classified ads that many local newspapers had relied upon. Indeed, Figure A1 (in Appendix A), which plots the evolution of U.S. newspapers' revenues over time, shows that revenues from classified ads declined earlier on and faster than other sources, and that this decline started precisely around the time of CL's expansion.

Our analysis exploits variation in the timing of the introduction of CL across U.S. counties. In addition, we also exploit differences in how heavily newspapers relied on revenues from classified ads at baseline, which captures how much they were affected by the entry of a new competitor in that market.

3. Data

Our analysis combines data on: i) Craigslist's expansion across the U.S.; ii) browsing behavior for a large panel of Internet users; iii) characteristics, organization and market outcomes of daily newspapers; iv) newspapers' content; v) survey data on media consumption; vi) political outcomes including voter turnout, split-vote and representatives' behavior in office.

3.1. Expansion timeline and Craigslist availability

Our analysis exploits the staggered introduction of CL across U.S. counties. To construct a measure of the availability of CL in each county, we first collect information on the timing of the entry of each of CL's current websites. For a subset of these, the information is directly available on the CL's "about" webpage.[9] For the others, we assigned the date of the first snapshot recorded by the Internet Archive.[10] Figure 1 shows the evolution of the number of CL websites between 1995 and 2010, along with the evolution of a proxy for Internet quality - the average number of Internet Service providers by zipcode. Figure A4 depicts the geographic distribution of CL websites in 2000, 2005, and 2010 respectively.

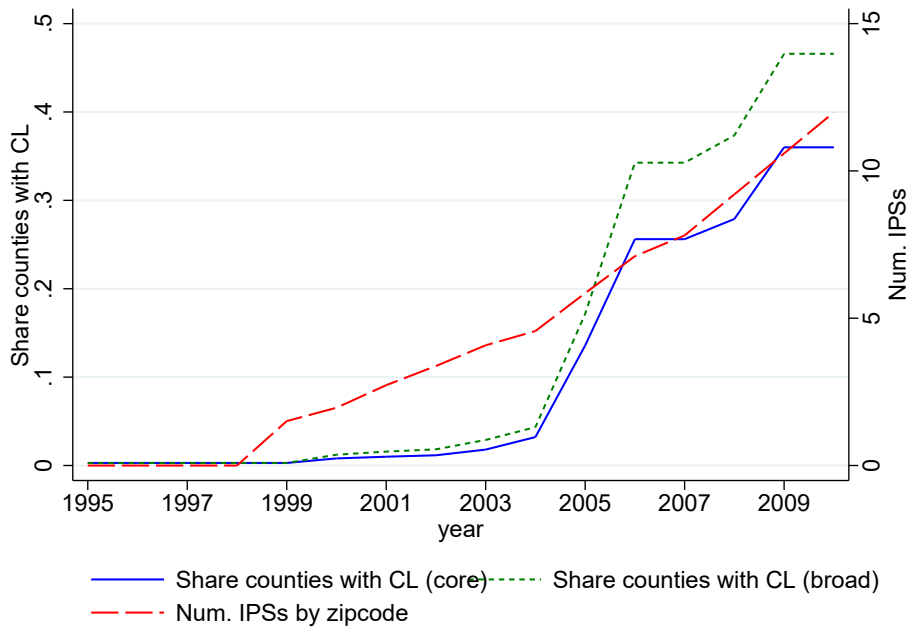
[7] A full list of the exceptions as of 2010, the end of our sample period, is available at: https://web.archive.org/web/20100706030043/https://www.craigslist.org/about/help/posting_fees

[8] For a profile of Craig Newmark and his business strategy see https://www.theguardian.com/technology/2006/feb/19/news.theobserver1.

[9] https://www.craigslist.org/about/expansion

[10] https://archive.org

FIGURE 1: CRAIGSLIST'S ROLL-OUT OVER TIME



Share of US counties with a local CL website (left axis) and average number of Internet service providers by zipcode (right axis).

Mapping websites to counties is not straightforward since the area actually served by the website may, in some cases, be larger than what the website's name indicates. To address this issue, we identify the relevant market to a CL website in two ways. In our baseline approach, we assume that CL websites serve primarily the county (or set of counties) containing the place indicated in the headline of the respective local website. The place is usually a single city or town, but can also be a combination of two or three nearby cities, a larger geographic area, or, in some cases, an entire state. We refer to this definition as the "core" CL market.

As an alternative, we identify all the counties that account for a non-negligible share of the ads posted on the website. To do so, we retrieve the snapshots of each website available from https://archive.org//, and code the exact location of all the ads posted on the first page of the "housing", "jobs", and "sales" sections. To get an accurate picture of the situation at the time the website started operating, we focus on the ads post in the first two years after the entry. We then match the resulting locations to a comprehensive list of towns, cities, and counties (if the location includes the word "county") in the same or a neighboring state. Finally, we consider all counties that account for at least 5% of the ads as part of what we define as the website's "broad" market.

Figure A5 depicts the geographic distribution of ads posted on https://brainerd.craigslist.org/

in the 1st and 2nd year after the opening of the website. In this case, the "core" market is represented by the central county (Crow Wing County) containing the city of Brainerd. This "core" county accounts for over 80% of total ads, while the "broad market" includes five additional neighboring counties. This is a typical pattern in our data: on average the "core" market accounts for 73% (median 76%) of posted ads once we exclude outliers.

3.2. Browsing data

To measure CL take-up over time, as well as the impact on newspapers' online editions, we also draw on data from the *Comscore*[11] panel, which tracks the browsing behavior of a large sample of US Internet users. The data have been shown to be representative of US online buyers (De los Santos et al., 2012) and cover the following years in our sample period: 2002, 2004, 2006 and 2007-2010. We aggregate the number of visits of the domain craigslist.org as well as total visits recorded by Comscore by county and year. Using information on the website corresponding to each newspaper in our sample (see next section for more detail), we similarly aggregate visits by newspaper website and year.

3.3. Newspaper characteristics and outcomes

The aim of our analysis is to shed light on how the entry of CL, and the subsequent decline in classified ads revenues, affected local US daily newspapers. To this end, we collect comprehensive data on a range of relevant variables including market outcomes, organization, and content. Our main source is a series of Newspaper Yearbooks published by Editor & Publisher for each year between 1995 and 2010 which we digitized using OCR software. The yearbooks contain detailed information for over 1,500 U.S. daily newspapers, including: address of the headquarters (HQ), circulation, and the list of staff members with names, broad job categories, and job titles. Figure A7 shows how the information is reported in the yearbooks for two example newspapers, one small and one large.

To understand whether and when a newspaper is affected by the entry of a CL website, we first need to match newspapers to counties. One way to do so is to assume that the newspaper market coincides with the county where its HQ is located. Since this a good approximation for the median newspaper, this approach is common in this literature (Gentzkow and Shapiro, 2010; Seamans and Zhu, 2014). We obtain the HQ county for each newspaper from the address reported in the E&P data.

An alternative and more precise method relies on the availability of detailed circulation data and consists in matching a newspaper to all counties where it is read in proportion to circulation. Zip-code-level circulation data are available from the Alliance for Audited

[11] https://wrds-www.wharton.upenn.edu/pages/about/data-vendors/comscore/

Media (AAM) for 2002, which we use to construct a weighted measure of CL availability by newspaper-year. The weights in this "broad" measure of CL availability are the fraction of the paper's subscribers in each county.[12] However, AAM data covers only about 40% of the papers in the E&P sample. For the newspapers for which no zip-code circulation data is available, we assign 100% of the circulation to the county where the paper's HQ is located.[13] Figure A6 shows the geographic distribution of circulation for a newspaper in our sample, the Brainerd Dispatch, with darker colors representing higher values. Crow Wing County where the newspaper's HQ is located, shown in blue, accounts for 81% of the total.[14] The median paper in our AAM data has about 85% of its total circulation in the headquarters county once we exclude outliers.

Whether and how much a newspaper is affected by the entry of CL likely depends on how heavily it originally relied on revenues from classified ads. To measure this baseline difference across papers, we use information on the presence in a newspaper's staff of one or more classified ads managers prior to CL's entry (i.e. in the year 2000), available from E&P. To validate this measure, we collect information from https://www.newspapers.com, an online newspaper archive, on the share of pages devoted to classified ads in a subset of 261 newspapers. The results, shown in Appendix B.1, indicate that, prior to the entry of CL into their market, newspapers with classified ad managers devoted a significantly larger page share to classified ads than other papers, about 20% more than non-classified manager papers with similar circulation.

3.4. Newspaper content

We are also interested in examining how changes in newspapers' organization triggered by the entry of CL influenced editorial decisions and news content. To this end, we use information from NewsBank, a database containing the text and metadata of all articles published in over 800 newspapers in our sample, beginning in 1999. We use the data in three ways. First, we extract a random sample of 2 million articles and apply a semi-supervised topic model to the text of the lead paragraph to assess changes in the the coverage devoted to various topics. Second, we perform keyword searches on the full text of all articles looking for names of specific politicians (e.g., "Rep. Paul Ryan", "Senator Dianne Feinstein", etc.), and use the number of mentions in a given newspaper/year as a complementary measure of

[12] We measure geographically disaggregated circulation only once, in 2002, and hence year-to-year variation is driven entirely by changes in CL availability and not by changes in circulation patterns.

[13] The papers that are missing from AAM are generally smaller papers and, if anything, less likely to have circulation beyond the county boundaries than the papers that appear in AAM. Papers which appear in AAM had median circulation in 2002 of 67K, compared to 14K for papers not appearing AAM. Hence, we believe that assigning all circulation to the headquarters county is a good approximation for these papers.

[14] Similarly to CL ads, we exclude outlier counties that account for less than 5% of total circulation.

the prominence of political issues, both national and local. Finally, to investigate how the entry of CL affects news coverage of corruption, we conduct targeted searches of corruption-related keywords in the full text of all articles in the database. These contain keywords for charges commonly prosecuted in public corruption cases against government officials, such as "bribery" or "conspiracy," as well as a list of eighteen phrases covering Hamilton's (2016) typology of principal-agent problems that can be exposed by investigative journalism. We aggregate counts of mentions by newspaper and year. Details of the procedure used to construct these variables are reported in Appendix B.2.

3.5. Survey data

To explore the impact of changes in newspapers' content on readers' news consumption habits, we use individual data from two large scale surveys. Our first source is the National Annenberg Electoral Survey (NAES), a nationally-representative rolling cross-sectional survey that was conducted in the lead-up to the 2000, 2004, and 2008 presidential elections. In particular, we use information on respondents' Internet access and self-reported media consumption in the week prior to the interview.

Our second source is the Survey of the American Consumer conducted by GfK Mediamark Research & Intelligence (GfK-MRI). The survey includes an extensive battery of questions about media consumption. In particular, we use information on respondents' self-reported readership of newspapers (in print and online, national and local), and news consumption on radio, TV, and on the Internet. We also exploit a question on what sections of the newspaper respondents usually read. We use data for the period 1999-2010.

3.6. Political outcomes

We are also interested in examining how the entry of CL, and the subsequent changes in newspapers organization and content influenced the behavior of voters and politicians. To this end, we collect data on several voting outcomes measured at county level, and on the behavior of members of Congress, measured at the Congressional district level.

For voters' behavior we look at electoral turnout, a standard measure of political participation and split-ticket vote, a measure of partisan polarization. We use county-level data on turnout and party vote shares in the House, Senate, and Presidential elections from the Congressional Quarterly (CQ) Voting and Elections Collection. This dataset covers all counties in the the US from 1992-2012. Split-ticket voting is based on county-level party vote shares and captures the extent to which voters support candidates from different parties in concurrent elections. Following Darr et al. (2018), we measure split voting as the absolute value of the difference between the Republican candidate vote share in the presidential election and

the Republican candidate vote share in House elections in the same county and year.[15]

Regarding politicians' behavior, we focus on House representatives and look at ideological extremism and party discipline, using data on roll-call votes available from Lewis et al. (2020). We focus on the years between 2003 and 2012, a period between decennial redistricting years during which district boundaries remain largely constant. To measure a politician's position on the liberal-conservative spectrum, we use information from the Americans for Democratic Action (ADA) annual scorecards, a source that has been widely used in the political science literature. The ADA score measures, for each representative in each roll-call vote, the degree of alignment with the group's (left-leaning) preferences on a scale from 0 (minimum alignment) to 1 (maximum alignment). Data are available from ADA for the period 2008-2012, and from Anderson and Habel (2009) for the previous years. Following Canes-Wrone et al. (2002), we compute ideological extremism as the ADA score of a given representative in a given year for Democratic members, and 1 minus the ADA score for Republican members, so that higher value capture positions closer to both ends of the spectrum.

Using information from roll-call votes we can also examine whether representatives became more likely to vote along party lines. To measure the degree of party discipline, we compute the share of roll-call votes in which a representative voted with the leadership of his/her party in any given year. This variable has been interpreted as a measure of how much a representative is responsive to party pressures as opposed to the preferences of her/his local constituency, with greater values reflecting lower relative weights on the latter (Snyder and Strömberg, 2010).

In addition, to measure the effort exerted by politicians in their legislative activity beyond roll-call voting, we collect information on the number of bills sponsored by each member of Congress in each year. These data are available for the entire sample period from the Policy Agendas Project (Adler and Wilkerson, 2020).

Finally, for our analysis of news coverage of corruption we are interested in controlling for the actual incidence of corruption at the local level. To this end, we use information on prosecutions of state and local government officials by the US Department of Justice (DOJ) available from the Transactional Records Access Clearinghouse (TRAC) database hosted by Syracuse University. The dataset contains 46,783 records of corruption cases from 1986-2019 in each of 94 federal court districts across the country. We compute the number of cases recorded in each year in each district, and match it with newspapers according to the

[15] Darr et al. (2018) use the Senate-President difference, but since only a third of Senate seats are contested in each election cycle, using House races expands the number of observations available. We measure vote shares for each office at the county level; House shares are computed by aggregating across all House votes cast by voters in the county (which for larger counties may be split across multiple representatives).

headquarter location.

3.7. Number of Internet Service Providers

To separate the impact of CL entry from a generic Internet effect, we control for a measure of the quality of local broadband Internet. In the absence of disaggregated data on Internet subscribers for this period, we follow Larcinese and Miner (2018) and Seamans and Zhu (2014) and use the number of Internet service providers (ISPs) registered by zipcode as a proxy.

These data are available for the period 1998-2008 from the Federal Communication Commission (FCC) and cover all providers with more than 250 high-speed lines in a state and transfer speed greater than 200 kilobits per second. We assign years before 1998 zero ISPs for all zip-codes, and use linear interpolation to fill missing data for years after 2008. We then aggregate the number of ISPs to the county level by taking the population-weighted average across zipcodes contained in a county.

This measure has been shown to be a strong predictor of the number of broadband subscribers at the state level, as well as at the county level in later periods when such disaggregated data are available (Larcinese and Miner, 2018). To further validate the number of ISPs as a proxy of local Internet penetration, we examine its correlation with self-reported Internet access from both the NAES and GfK-MRI surveys. Figure A8 confirms a strong positive relationship.

3.8. Other county characteristics

Throughout our analysis we also use data on the following variables: population (from the National Center for Health Statistics), income per capita, share of the population in urban areas, share of the population with college education, share of the population who rent, racial composition and median age (all from the 2000 Census), as well as unemployment rate (from the Bureau of Labor Statistics). We obtain electoral data including turnout and Republican vote share in presidential and Congressional elections from the David Leip Election Atlas.

4. Empirical Strategy

4.1. Determinants of CL entry

To implement this strategy, it is necessary first to understand what factors determine the timing of CL entry in a locality. Anecdotal evidence suggests that CL prioritized larger markets and areas with good access to broadband Internet, which was crucial for the user to take full advantage of the platform. Importantly for our purposes, CL is not in the news

business, and thus is not likely to have considered demand-side factors in the news market in determining where to enter. Also importantly, CL is privately held and has always operated as a mixture of profit-making business and community service. The fact that CL did not maximize profits[16] allowed more flexibility for the idiosyncracies of its influential founder and its early user base to determine the timing of the rollout, rather than systematic factors in the underlying local news markets.

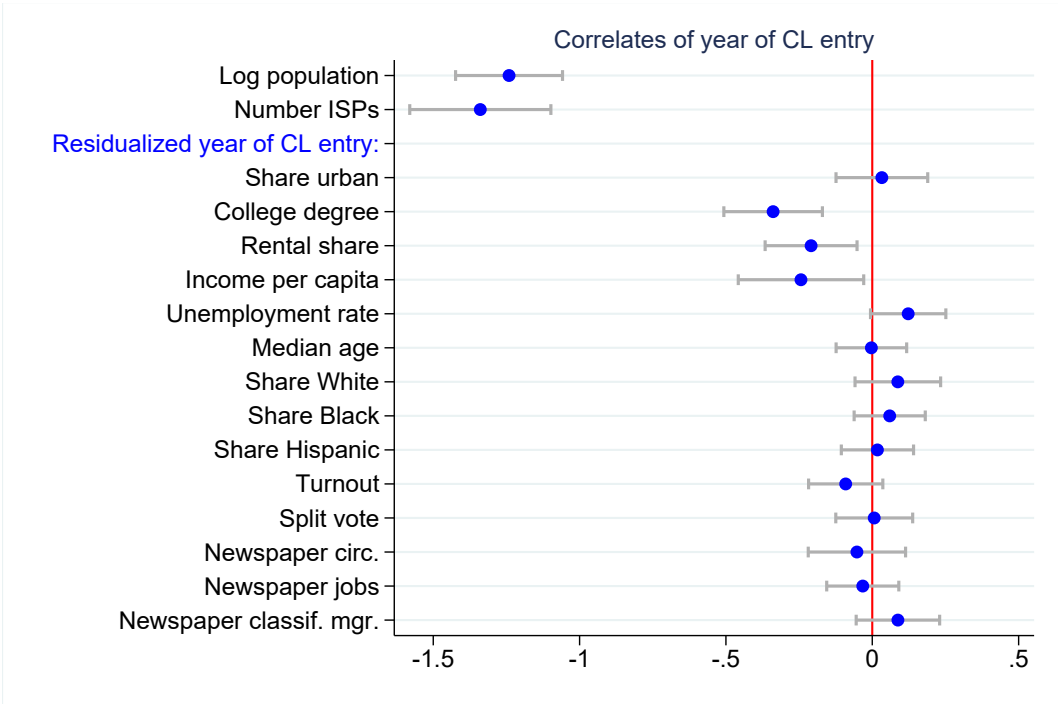
To validate the assumption of conditional exogeneity of CL's timing of entry, in Figure 2 we examine the set of counties with newspaper HQs in 2000 and plot the correlations of the year of CL's entry into a given county with its cross-sectional characteristics. The first two coefficients confirm that, as suggested by anecdotal accounts, population and the quality of the local Internet connection, as measured by number of ISPs, were important considerations in CL's decision on which cities to enter first. The magnitudes of both coefficients are sizable: a one-standard deviation higher population (number of ISPs) is associated with CL entering a county about 15 months (16 months) earlier. Together, these two variables account for 37% of the variation in year of entry. Given this strong relationship, our analysis will control for log population and number of ISPs throughout.

The other coefficients are from separate univariate regressions of the time of CL's entry residualized by log population and number of ISPs. The figure suggests that, conditional on these two variables, several demographic and political variables, including share urban population, racial composition, unemployment rate, age, electoral turnout and split-vote, appear unrelated to the timing of CL's entry. There are three exceptions: consistent with the profile of CL's early user base, counties with higher share of college educated population, higher rental share of housing and higher income per capita experience significantly earlier entry (3 to 5 months per standard deviation). We therefore also consider specifications controlling for each of the variables listed above.

Finally, and crucially for the purpose of our analysis, we find no relationship between timing of entry and the state of local newspapers as measured by circulation, number of jobs or the presence of classified managers. In other words, we find no evidence of CL targeting particular newspaper characteristics in the decision on where to enter first.

[16] The judge in a 2010 civil action against Newmark by eBay concluded that "Craigslist does not expend any great effort seeking to maximize its profits or to monitor its competition or its market share." (eBay Domestic Holdings Inc. v. Newmark, Delaware Court of Chancery Civil Action No. 3705-CC, decision dated 2010-09-09, https://h2o.law.harvard.edu/cases/3472.) The absence of profit maximization motivated the suit, as eBay (which held a substantial stake in CL) argued that founder Craig Newmark had failed his fiduciary duty to maximize returns to shareholders.

FIGURE 2: CORRELATES OF CL ENTRY



Coefficients and 95% confidence intervals representing the univariate correlations of (residualized) year of CL entry with various county characteristics. All variables are measured in the year 2000 and standardized to facilitate the comparison of magnitudes. The sample consists of all counties with newspaper HQs in the year 2000.

4.2. Newspaper-level regressions

To estimate the effect of CL entry on newspaper-level outcomes, we employ a difference-in-differences strategy exploiting CL's staggered introduction across US counties, combined with differences across newspapers in ex-ante reliance on classified advertising. The sample consists of the all newspapers covered by E&P, excluding ones with national circulation - i.e. the New York Times, USA Today and the Wall Street Journal.

The following equations summarize our approach:

$$Outcome_{nct} = \beta PostCL_{ct} + \phi_n + \psi_t + \rho' X_{ct} + \upsilon_t' Z_{c0} + \varepsilon_{nct}, \tag{1}$$

$$Outcome_{nct} = \beta PostCL_{ct} + \gamma PostCL_{ct} \times ClassifMgr_{n0} + \phi_n + \psi_t + \rho' X_{ct} + \upsilon_t' Z_{c0} + \varepsilon_{nct} \tag{2}$$

$Outcome_{nct}$ is one of the outcomes of interest for newspaper n, headquartered in county c, at time t. $PostCL_{ct}$ is a an indicator variable equal to one for years after the entry of CL in county c, following the "core" definition of CL markets (based on location indicated in the website name) and newspaper markets (based on newspaper HQ).[17] ϕ_n and ψ_t are newspaper and year fixed effects, respectively. The vector X_{ct} includes controls for population and number of Internet providers, and the vector Z_c0 includes additional county-level controls from the 2000 census which we interact with year fixed effects. In alternative specifications we also control for state $\times$ year fixed effects or for DMA $\times$ year fixed effects thus restricting the comparison to newspapers operating in the same state or in the same media market. Finally, $ClassifMgr_{n0}$ is an indicator variable equal to one if the newspaper had a classified manager at baseline. We cluster standard errors by the area affected by the entry of a given CL website (i.e., a single county or group of counties), or, for newspapers never affected by CL, by county.

Equation 1 estimates the impact of CL's entry under the assumption that the timing of entry is conditionally uncorrelated with pre-existing trends in these outcomes. To evaluate the plausibility of this assumption, for each outcome we we will also estimate an event-study version of specification 1. In light of recent work showing that two-way fixed effect estimates can be biased in settings where treatments effects are heterogeneous over time or across groups, we present event-studies based on the time-corrected Wald estimator proposed by de Chaisemartin and D'Haultfoeuille (2020).

[17] We obtain similar results - reported in Appendix E - using the respective "broad" definitions. In that case $post_CL$ is defined as 1 after the entry of CL in any county contained in the .

4.3. County-level regressions

To estimate the impact of CL entry on outcomes measured at the county level, we estimate versions of equations 1 and 2 aggregated by county, i.e.:

$$Outcome_{ct} = \beta PostCL_{ct} + \phi_c + \psi_t + \rho' X_{ct} + \upsilon_t' Z_{c0} + \varepsilon_{ct}, \tag{3}$$

$$Outcome_{ct} = \beta PostCL_{ct} + \gamma PostCL_{ct} \times ClassifMgr_{c0} + \phi_c + \psi_t + \rho' X_{ct} + \upsilon_t' Z_{c0} + \varepsilon_{ct} \tag{4}$$

When estimating this specification we focus on the sample of all counties where at least one newspaper was located at baseline, and compute $ClassifMgr_{c0}$ as the circulation-weighted average across all newspapers HQ-ed in county c.

4.4. Congressional district-level regressions

When we examine how the behavior of representatives responded to the CL shock, we run into the problem that there is not a one-to-one correspondence between newspapers and congressional districts. We thus require some measure of the importance of each newspaper to each congressional district, to allow us to weight up changes in newspapers' operations to the congressional district level. Snyder and Strömberg (2010) show that *ReaderShare* - i.e., the fraction of a newspaper's readership that resides in a given district - is a good predictor of the amount of coverage that a paper devotes to the representative from that district.[18] They further show that *Congruence*, the product of *ReaderShare* with the paper's market share in a district, is a good measure of the paper's overall importance to the behavior of the district's representative. We will therefore run regressions at the newspaper-district level, interacting the Craigslist-entry measure with the reader share of the newspaper in a given district:

$$Outcome_{dt} = \beta PostCL_{nt} + \gamma PostCL_{nt} \times ReaderShare_{nd} + \phi_d + \psi_t + \rho' X_{dt} + \upsilon_t' Z_{d0} + \varepsilon_{dt}, \tag{5}$$

$$ReaderShare_{nd} = Circulation_{nd} / \sum_{d'} Circulation_{nd'} \tag{6}$$

where $ReaderShare_{nd}$ are computed for each pair of district and newspaper using pre-Craigslist circulation data, using the method described in Snyder and Strömberg (2010).[19]

[18] Our coverage data confirms that there is a strong relationship between *ReaderShare* and coverage of a representative.

[19] For papers with data in AAM, we use zipcode-level circulation to match readers to districts. For papers that do not appear in AAM, we match to districts that overlap the county, and use zip-level population shares in place of subscriber shares to allocate subscribers. These shares are computed once, using pre-Craigslist entry circulation data, implying that there is no over-time variation in our measure of reader share or market share.

We weight each observation in the regression by the paper's market share in the district ($MarketShare_{nd} = Circulation_{nd} / \sum_{n'} Circulation_{n'd}$) such that each district-year observation gets total weight of one, with more weight given to papers with high market share in the district. The product of *MarketShare* and *ReaderShare* is Snyder and Stromberg's definition of "congruence," and hence our estimates of γ in Equation (5) give heaviest weight to changes in newspaper-district pairs with high congruence. The underlying logic is that we expect changes in the content of papers to matter more for districts with which the paper has high congruence.

5. Results

5.1. Craigslist Take-up

We first confirm that Internet users in a given area were more likely to visit CL's URL following the opening of a local CL website. To this end we use the data on web traffic from Comscore described in section 3. In Table 1 we estimate various versions of our baseline specification (equation 1) using as dependent variable the number of visits to the CL website recorded in a county in a given year. In all regressions we control for the total number of visits to any website recorded by Comscore in the same county-year, a measure of overall online activity. The results indicate that, after the entry of a local CL website, the number of visits to CL's URL increases significantly. The result is robust to expanding the set of county controls (col. 2), to controlling for state $\times$ year fixed effects (col. 3), and even for DMA $\times$ year fixed (col. 4), that is restricting the comparison to different counties within the same media market.

To get a sense of the timing of CL's take-up, in Figure 3 we plot the coefficients on four lags and four leads relative to the entry, estimated from a single regression. The graph indicates no clear pre-trend, and suggests that the effect was immediate but further intensified over the following years. This pattern is consistent with network effects in the adoption of the website, since a larger number of users (and hence, a higher volume of ads) increases the value of the platform.

5.2. Number of newspapers

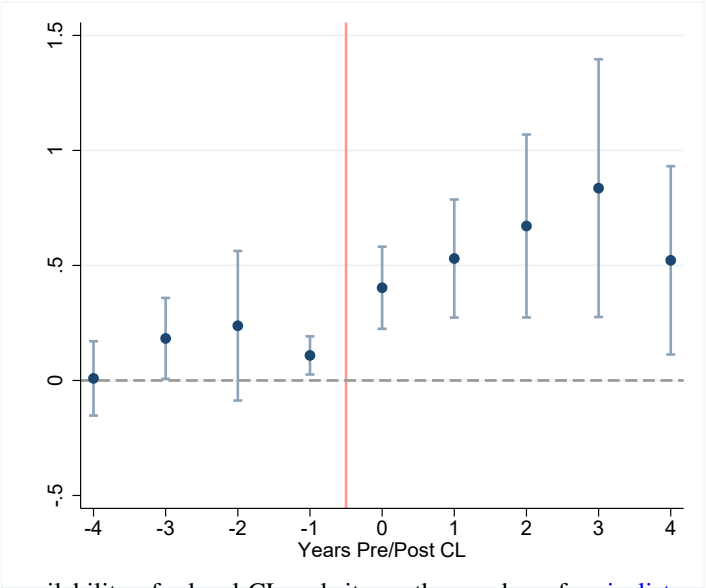
Next we examine how the entry of CL in a given market affected the number of active newspapers. Indeed, one possibility is that the drop in revenues due to competition from CL might have been so extreme as to force some local papers to close or merge. In Table 2 we estimate our baseline specifications using as dependent variable the number of newspapers that,

TABLE 1: CRAIGSLIST.ORG VISITS

	Dependent variable: CL visits (IHS)			
	(1)	(2)	(3)	(4)
Post-CL	0.456***	0.175**	0.217***	0.178***
	(0.132)	(0.082)	(0.074)	(0.059)
Total Comscore visits (IHS)	Yes	Yes	Yes	Yes
Log population, # ISPs	Yes	Yes	Yes	Yes
2000 county char. × Year FEs	No	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes
Observations	20396	20396	20396	20007
Number of counties	3043	3043	3043	2986
R^2	0.79	0.82	0.83	0.85
Mean dependent variable	2.36	2.36	2.36	2.34

Regressions of number of visits of the domain craigslist.org by county and year (IHS-transformed) on an indicator for the availability of a local Craigslist website. All specifications control for total visits recorded in Comscore (IHS-transformed). County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE 3: CRAIGSLIST.ORG VISITS – EVENT STUDY



Dynamic effect of the availability of a local CL website on the number of craigslist.org visits by county and year (IHS-transformed). Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include total Comscore visits (IHS transformed), log population and number of Internet service providers. Standard errors clustered by CL-area.

according to the E&P data, were operating in a given county in a given year. In this case our unit of analysis is a county-year, and the sample includes all counties where at least one newspaper was headquartered in 2000 (i.e., before CL). We find no evidence that CL entry caused a drop in the number of newspapers in a county. If anything, the dummy *Post-CL* displays a positive and significant coefficient in column 1, which however becomes insignificant when the set of controls and fixed effects is expanded. CL does not seem to affect the number of newspapers even in those counties where newspapers relied more heavily on classified ads *ex ante*; indeed, in columns 5 through 8, the coefficient on the interaction between *Post-CL* and *Classif. Manager* is small and never statistically significant. The same picture emerges when looking at the event-study graph depicted in Figure 4: all the coefficients for the years after CL entry are close to zero and statistically insignificant.[20] Taken together, these findings do not support the view that CL contributed to the disappearance of local newspapers. Yet, it is possible that lower advertising revenues affected the organization, functioning, and content of newspapers in other ways, a hypothesis which we explore below.

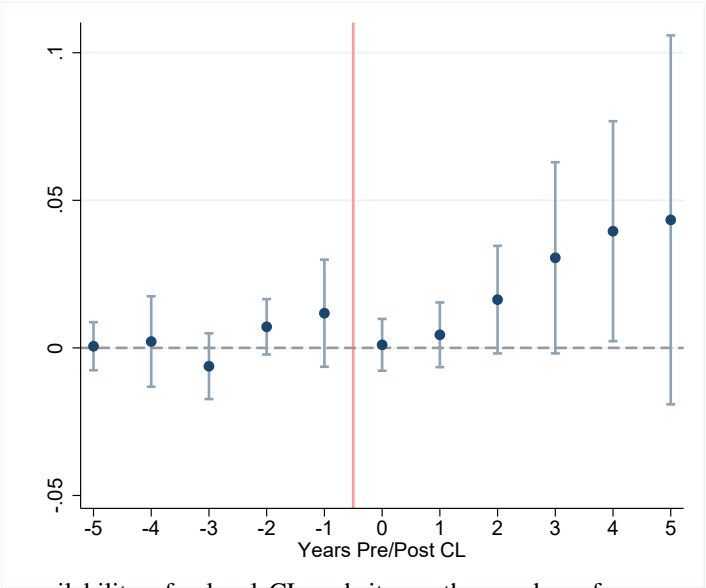
[20] Given the lack of evidence that CL affected newspapers' exit and entry, in the remainder of our analysis we use the full unbalanced panel as baseline. We obtain similar results for the balanced panel of newspapers that remained in the sample for the entire period of analysis, which we report in Appendix **??**.

TABLE 2: NUMBER OF NEWSPAPERS

	Dependent variable: Number of newspapers HQ-ed in county							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	0.025**	0.019	0.009	0.010	0.033**	0.022*	0.015	0.009
	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.015)
Post-CL × Classified Mgr.					-0.014	-0.006	-0.010	0.002
					(0.019)	(0.019)	(0.019)	(0.020)
Log population, # ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × Year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	19424	19424	19424	18672	19360	19360	19360	18608
Number of counties	1214	1214	1214	1167	1210	1210	1210	1163
R^2	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Mean dependent variable	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18

Regressions of the number of newspapers HQ-ed by county and year on an indicator for the availability of a local Craigslist website and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE 4: NUMBER OF NEWSPAPERS – EVENT STUDY



Dynamic effect of the availability of a local CL website on the number of newspapers HQ-ed by county and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

5.3. NEWSPAPER STAFF SIZE AND COMPOSITION

We start by examining to what extent the entry of CL local websites influenced the organization of local newspapers. The first outcome we focus on is staff size. In Table 3 we estimate our baseline specifications using as dependent variable the number of jobs reported in the E&P yearbooks for each newspaper in a given year. In the first four columns we examine the effect of being in the *Post-CL* period. Results in columns 1-4 indicate that following the entry of CL, the number of jobs in local papers decreased significantly. The effect amounts to about 5% of the mean, and is robust to the inclusion of additional controls and finer fixed effects. The results in the last four columns confirm that the effect is mainly driven by newspapers that relied more on classified ads prior to CL, for which the effect, looking at the most demanding specification (column 8), is over 12% relative to the mean. Figure 5 displays the associated event-study graph, which shows the absence of any significant pre-trend, and that the effect, which is significant in the first year, gets considerably larger as the years pass.

This result is robust to several additional checks. First, we verify that the indicator for classified manager does not simply capture newspaper size as in our data newspapers with larger circulation, are more likely to have a dedicated classified manager (correlation = 0.26). In Table C1 we show that the results are robust to controlling for baseline circulation interacted with $post_C L$, and that this interaction has no significant effect on jobs-count.

Second, we consider as dependent variable the number of (unique) employees instead of the number of job titles. These two variables can differ if, for example, financial difficulties push a newspaper to assign to the same person multiple jobs that were previously carried out by different people. Table C2 and Figure C1 show that we obtain very similar results.

One important question that our data allow us to examine is what categories of workers were mostly affected by staff cuts. In fact, based on the job title we can determine whether a worker holds a managerial or an editorial position, and, for editorial staff, we can identify the corresponding topical area (e.g., politics, sports, entertainment, etc.).[21]

In the first two columns of Table 4, we estimate our main specification with the *Post-CL*× *Classif. Manager* interaction separately for managerial staff and editorial staff. The results indicate that, for the newspapers most affected by the entry of CL, staff cuts concerned both types of positions, with the effect on managerial positions being larger (i.e., 19% vs. 7% of the mean). In the following columns we examine what topical areas were most affected by cuts in editorial staff, looking in particular at politics, sports, and entertainment. The evidence indicates that newspapers most affected by CL were significantly less likely to have dedicated political editors and reporters after CL's entry, while the same was not the case for sports or entertainment. One interpretation of this result is that, when facing financial difficulties, newspapers affected by CL opted to cut staff especially in areas - like local politics - that readers do not value as much and for which producing quality content is more costly.

On the other hand, despite this significant downsizing, we find no effect on the number of pages published by affected newspapers (Appendix Table C3 and Figure C2). Together, these results imply increased workload per staff member and may have implication for the distribution of editorial priorities. We examine this issue further in the next section where we look at how the entry of CL affected the evolution of news content.

5.4. Newspaper content

Next we explore how the transformations in newspapers' organization documented above translate into changes in news content, in general, and news coverage of local politics, in particular. This is a question of paramount importance considering that, for most citizens, local newspapers still represent the main source of political information which allows them

[21] Specifically, we classify as managerial positions the jobs listed in the E&P section "Corporate/ General Management", and as editorial positions the jobs listed in the sections "News Executives" and "Editorial Management" (see figure A7). Regarding topical areas, we classify jobs containing the keyword "sports" as sports related, ones containing the keywords "entertainment"/ "movie" / "music" / "theatre"/ "travel" / "lifestyle" as entertainment-related, and ones containing the keywords "politics", "government", "Washington" as politics-related.

TABLE 3: NUMBER OF JOBS

	Dependent variable: Newspaper number of jobs							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	-2.030***	-1.084***	-1.242***	-1.262***	-0.103	0.453	0.274	0.163
	(0.398)	(0.390)	(0.418)	(0.404)	(0.412)	(0.411)	(0.432)	(0.442)
Post-CL × Classified Mgr.					-3.623***	-2.975***	-2.911***	-2.802***
					(0.575)	(0.548)	(0.549)	(0.594)
Log population, #ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	22724	22724	22723	21968	22516	22516	22515	21762
Number of newspapers	1540	1540	1540	1492	1505	1505	1505	1459
R^2	0.90	0.91	0.91	0.92	0.90	0.91	0.91	0.92
Mean dependent variable	21.31	21.31	21.31	21.14	21.38	21.38	21.39	21.23

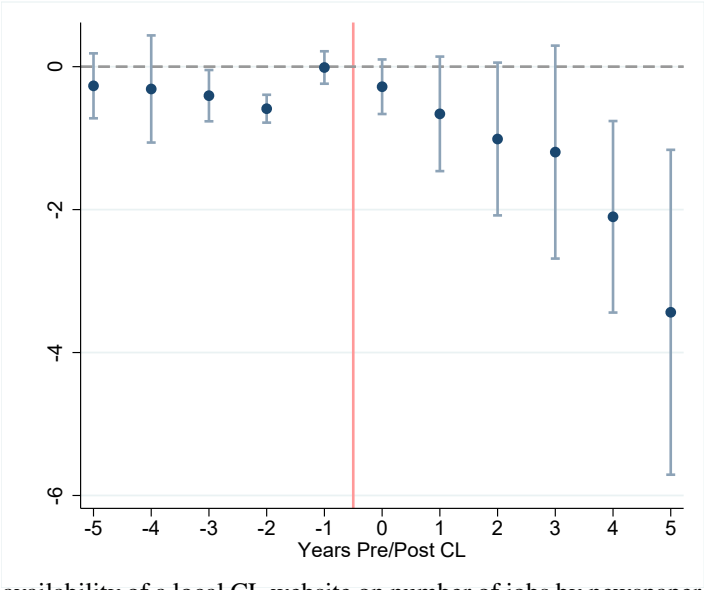
Regressions of number of jobs by newspaper and year on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 4: JOBS BY TYPE AND TOPIC

	(1) Num. Managers	(2) Num. Editors	(3) Editor Politics	(4) Editor Sports	(5) Editor Entertainment
Post-CL	0.050	0.138	0.022	-0.019	0.041*
	(0.086)	(0.306)	(0.019)	(0.018)	(0.024)
Post-CL × Classified Mgr.	-0.667***	-0.782*	-0.050**	-0.003	-0.016
	(0.122)	(0.444)	(0.024)	(0.021)	(0.029)
Log population, #ISPs	Yes	Yes	Yes	Yes	Yes
2000 county char. × Year FEs	Yes	Yes	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes
Observations	21834	22155	21454	21558	21445
Number of newspapers	1503	1504	1504	1504	1504
R^2	0.80	0.89	0.69	0.51	0.57
Mean dependent variable	3.46	10.60	0.40	0.84	0.33

Regressions of number of jobs by type and indicators for the presence of an editor dedicated to a specific topic on an indicator the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE 5: NUMBER OF JOBS – EVENT STUDY



Dynamic effect of the availability of a local CL website on number of jobs by newspaper and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

to monitor and keep politicians accountable (Mahone et al., 2019; Snyder and Strömberg, 2010). To analyse changes in news content we apply text analysis techniques to large corpora of news articles published in a subset of the newspapers in our original sample. Our analysis is composed of three parts.

First, we study how competition from CL affected the volume of coverage devoted to different topics. To do so, we estimate a topic model on a corpus consisting of the first paragraphs of over 2 million randomly drawn articles published on over 800 newspapers between 2000 and 2010, available from Newslibrary.com. We use a particular type of topic model - the Correlation Explanation (CorEx) (Gallagher et al., 2017a). This model has the advantage that it tends to produce coherent topics for corpora consisting of short texts, and it furthermore which allows us to define anchor words to generate topics corresponding to specific areas. Specifically, we seed the model to produce, among others, topics associated with: i) President, ii) Congress, iii) local politics, and iv) foreign policy.[22] Figure B1 reports the most frequent words associated with each of the ten topics generated by the algorithm, with the anchor words underlined. For each article in the corpus, we obtain a probability associated with each of the 10 topics, and aggregate the probability weights by newspaper and year. In Table 5 we estimate our basic specifications using as dependent variables the average

[22] Additional details about the procedure are reported in Appendix B.2.

weights on each of four political topics. The results indicate a general decline in news coverage of politics. For newspapers most affected by CL, the drop is significant for presidential, congressional, and foreign politics, and marginally insignificant for local politics (which, importantly, does not include coverage of local representatives or senators). Figure 6 shows the event-study graph for congressional politics, which documents the absence of any pre-trend, and a significant, rapid, and persistent decline. In Table C5 we replicate the exercise for the remaining un-anchored topics (i.e., accidents, events, traffic, obituaries, sports, crime). Results are mixed, but for none of the categories is the coefficient on the interaction *Post-CL* $\times$ *Classif. Manager* significant.

TABLE 5: TOPIC MODEL: WEIGHT ON POLITICAL TOPICS

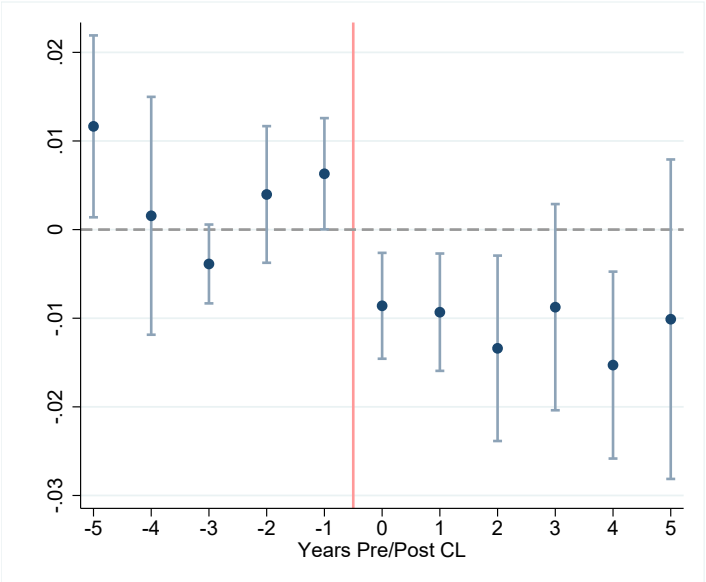
	(1) president	(2) local	(3) congress	(4) foreign	(5) president	(6) local	(7) congress	(8) foreign
Post-CL	-0.010**	-0.010*	-0.006**	-0.001	-0.001	-0.003	-0.000	0.005
	(0.004)	(0.006)	(0.003)	(0.003)	(0.005)	(0.007)	(0.003)	(0.004)
Post-CL $\times$ Classified Mgr.					-0.017***	-0.013	-0.011**	-0.011**
					(0.006)	(0.008)	(0.005)	(0.004)
Full county controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7178	7178	7178	7178	7129	7129	7129	7129
Number of newspapers	869	869	869	869	862	862	862	862
R^2	0.52	0.47	0.41	0.54	0.52	0.47	0.40	0.53
Mean dependent variable	0.21	0.31	0.10	0.10	0.21	0.31	0.10	0.10

Regressions of the average probability of an article covering a particular topic by newspaper and year on an indicator for the availability of a local Craigslist website, and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

To further explore changes in political coverage, we then examine how frequently local newspapers report about local and national politicians. Based on keyword searches on Newsbank, we identify all articles published in a given newspaper in a given year that contain the names of House representatives elected in the district corresponding to the newspaper HQ, and the name of national politicians, respectively.[23] In Table 6 we estimate our main specification using as dependent variable the number of articles mentioning local representatives (as a share of total articles). We find that, following the entry of CL, coverage of local politicians declines, particularly for newspapers that relied more heavily on classified ads, for which the drop is 45% of the sample mean. Again, the corresponding event study graph, reported in

[23] Details of the keyword searches are reported in Appendix B.2.

FIGURE 6: CONGRESSIONAL POLITICS TOPIC - EVENT STUDY



Dynamic effect of the availability of a local CL website on the average probability of a Congressional politics topic by newspaper and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

Figure 7, confirms the absence of pre-trends, and that the effect, already significant in the first year, builds up with time. In Table C6 we perform the same exercise for national party leaders: the President and leadership of both parties in each chamber of Congress. We find a small negative effect overall, but the interaction term with classified manager is positive and similarly sized, indicating that the most CL-affected papers had little change in coverage of national politicians. This difference indicates that the observed coverage declines for local representatives derived from cuts to reporting staff producing local politics coverage as opposed to a general decline in attention to politics.

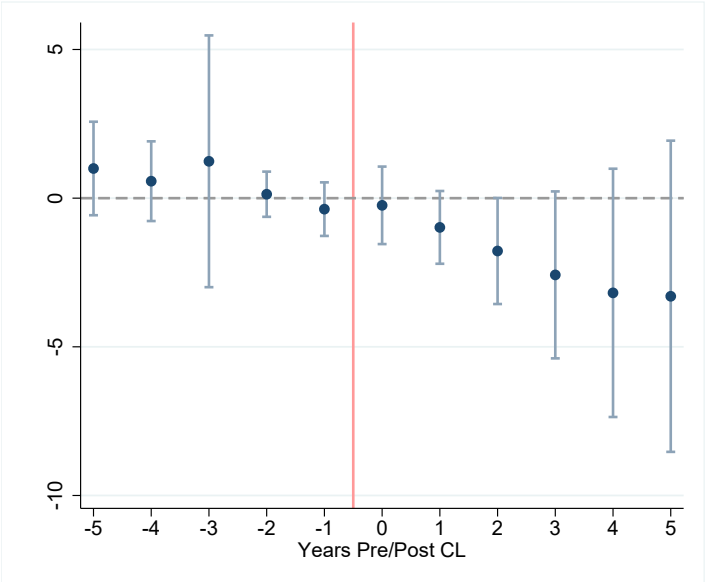
Finally, we examine to what extent competition by CL affects the ability of local newspapers to inform citizens about the (mis)conduct of local officials. To this end, as explained in section 3 and in greater detail in Appendix B.2, for each newspaper and year we identify in the Newsbank database all articles that include terms associated with public corruption (either legal charges or phrases from the Hamilton's taxonomy). Using this measure in Table 7 we test the impact of CL on news coverage of corruption. Crucially, in all specifications we control for the number of public corruption cases brought in the federal court district where the newspaper is based, since we aim to measure the effect on public scrutiny conditional on the underlying level of corruption. The results indicate that, regardless of the metric and the specification, coverage of corruption decreased significantly in newspapers affected by the

TABLE 6: MENTIONS OF LOCAL CONGRESSIONAL REPRESENTATIVES

	Dependent variable: Articles mentioning House representative per 1000							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	-0.303	-0.605**	-0.473	-1.286**	0.523	0.205	0.737	-0.068
	(0.302)	(0.306)	(0.394)	(0.592)	(0.534)	(0.517)	(0.555)	(0.611)
Post-CL × Classified Mgr.					-1.177**	-1.118**	-1.638***	-1.599**
					(0.568)	(0.551)	(0.594)	(0.674)
Log population, # ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × Year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	2089	2089	2033	1440	2081	2081	2025	1424
Number of newspapers	286	286	280	206	285	285	279	204
R^2	0.19	0.19	0.20	0.30	0.19	0.19	0.20	0.30
Mean dependent variable	3.69	3.69	3.73	3.78	3.69	3.69	3.74	3.78

Regressions of the share of articles mentioning the name of a local House representative by newspaper and year on an indicator for the availability of a local Craigslist website, and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE 7: MENTIONS OF LOCAL CONGRESSIONAL REPRESENTATIVES - EVENT STUDY



Dynamic effect of the availability of a local CL website on the share of articles mentioning the name of a local House representative by newspaper and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

entry of CL.

5.5. NEWS CONSUMPTION

How did readers react to the changes newspaper organization and content documented above? One way to tackle this question is by looking at the evolution of the total number of copies sold, i.e., newspaper circulation, available from E&P for all the newspaper in our sample. In Table 8 we estimate our standard set of regressions using as dependent variable yearly circulation. The coefficient on $Post - CL$ is negative although it loses significance when including state or DMA fixed effects interacted with year dummies. The effect is instead negative, large, and highly significant for the newspapers most affected by CL. The size of the coefficient is fairly stable across specifications and when all county controls are included, with magnitude of about 9% of the sample mean. The corresponding event study graph, reported in Figure 8, shows the absence of pre-existing trends and a steady increase in the effect within a few years of CL's entry.

One crucial question is to what extent the lower readership of local newspapers associated with the entry of CL is offset by increased consumption of other news sources, both offline and online. Indeed, if for instance readers merely substitute newspapers for other sources that cover similar content, such mode-switching, though detrimental for the papers, would

TABLE 7: COVERAGE OF PUBLIC CORRUPTION CRIMES AND CONFLICTS OF INTEREST

	Charges mentions (IHS)			Hamilton mentions (IHS)		
	(1)	(2)	(3)	(4)	(5)	(6)
Post-CL	0.157*	0.070	0.034	0.142**	0.072	0.042
	(0.081)	(0.080)	(0.088)	(0.069)	(0.070)	(0.073)
Post-CL × Classif. Mgr.	-0.294***	-0.229**	-0.278***	-0.214**	-0.160*	-0.210**
	(0.100)	(0.097)	(0.101)	(0.089)	(0.087)	(0.089)
County controls #1	Yes	Yes	Yes	Yes	Yes	Yes
County controls #2	No	Yes	Yes	No	Yes	Yes
Corruption cases	Yes	Yes	Yes	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	No	Yes	Yes	No
State×Year FEs	No	No	Yes	No	No	Yes
Observations	6,077	6,042	6,042	6,077	6,042	6,042
Number of newspapers	800	794	794	800	794	794
R^2	0.77	0.78	0.80	0.77	0.77	0.80
Mean dependent variable	5.7	5.7	5.7	6.7	6.7	6.7

OLS regressions in all columns. Standard errors clustered by CL-area (broad definition). Dependent variable is the hyperbolic arcsine of the yearly count of mentions of key words corresponding to commonly prosecuted public corruption cases (columns 1-3) or key words in the conflict-of-interest typology introduced by Hamilton (2016). See Appendix B.2 for details.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

not necessarily imply a reduction in actual news consumption. To shed light on this aspect, we use the survey data on news consumption from NAES and GfK described in section 3.

In Table 9 we test the effect of CL on the self-reported probability (or frequency) of reading a newspaper using individual level regressions. The surveys include information on the respondent's county of residence but only limited information on the specific newspaper she reads.[24]. Nonetheless, we differentiate between respondents who report most frequently reading a national newspaper, i.e. the *New York Times*, USA Today or *the Wall Street Journal*, and the rest which we presume are more likely exposed to local papers.

With this distinction, we rely on the county of the respondent and re-define the variable $ClassifMgr$ as the circulation-weighted average across newspapers based in that county. In all regressions we also control for a range of respondent's characteristics including age, race and education. The results based on the NAES data - reported in columns 1-4 - document a significant decline in both the probability and the frequency of reading non-national newspapers for people in areas where newspapers were most affected by CL. The effect is quite sizeable, corresponding to about 7% of the sample mean for both outcomes. Results are qualita-

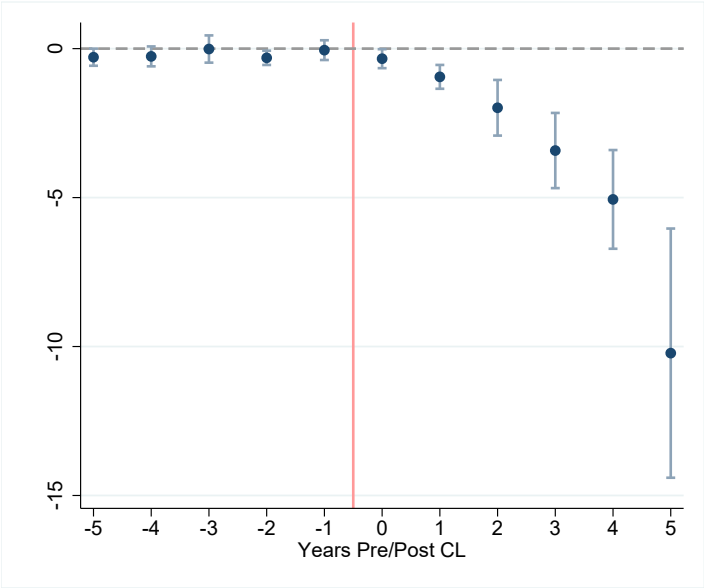
[24] Respondents in both surveys are asked which newspaper they read most often but this question has a large non-response rate - XXXXX in NAES and XXXXX in Gfk.

TABLE 8: CIRCULATION

	Dependent variable: Newspaper circulation in thousands							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	-2.712***	-1.468**	-1.194	-0.892	0.012	0.108	0.397	0.700
	(0.599)	(0.645)	(0.765)	(0.939)	(0.752)	(0.874)	(0.953)	(1.231)
Post-CL × Classified Mgr.					-5.148***	-2.991***	-2.896***	-2.921**
					(1.297)	(1.096)	(1.055)	(1.351)
Log population, #ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	22950	22950	22948	22195	22633	22633	22631	21880
Number of newspapers	1555	1555	1555	1507	1506	1506	1506	1460
R^2	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.99
Mean dependent variable	33.34	33.34	33.34	32.89	33.40	33.40	33.39	32.93

Regressions of circulation by newspaper and year, measured in thousands of copies, on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE 8: CIRCULATION – EVENT STUDY



Dynamic effect of the availability of a local CL website on circulation (in thousands of copies) by newspaper and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

tively similar, though somewhat smaller in size ($\approx 4\%$), $usingtheGfKdata.Sinceinbothsurveysrespondentsarea$ Both NAES and the GfK survey include questions about news consumption via sources other than local newspapers which we use to study substitution patterns. The results are reported in Tables C8 and C9. In columns 1 and 2 we examine how CL affected the readership of national newspapers. In both surveys we find that the entry of CL in an area is associated with an increase in the likelihood that local respondents read national papers. Yet, the effect does not appear to be stronger in areas where newspapers relied more heavily on classified ads. The remaining columns show no significant effect on news consumption on TV, radio, and online, including in areas where newspapers were most affected by CL.[25]

Taken together these findings suggest that the lower readership of local papers associated with the entry of CL was only partially compensated by increased news consumption via other means, and is hence likely to translate in a net decrease in political awareness.

5.6. POSSIBLE MECHANISMS

The decline in readership documented above is consistent with at least two explanations. First, it is possible that newspapers respond to the shock to classified ad revenues by increas-

[25] The analysis of online news consumption with NAES data is not possible since the question is not asked consistently across waves.

TABLE 9: SELF-REPORTED NEWSPAPER READERSHIP

	NAES				GfK-MRI	
	Read newspaper Dummy		Read newspaper Days per wk		Read newspaper Dummy	
	(1)	(2)	(3)	(4)	(5)	(6)
Post-CL	-0.023	0.019	-0.108	0.118	-0.016***	-0.003
	(0.017)	(0.015)	(0.090)	(0.108)	(0.005)	(0.008)
Post-CL × Classified Mgr.		-0.053**		-0.284*		-0.018*
		(0.164)		(0.027)		(0.009)
Respondent controls	Yes	Yes	Yes	Yes	Yes	Yes
County controls	Yes	Yes	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	108332	107503	108332	107503	253513	251442
Number of counties	1207	1203	1207	1203		
R^2	0.06	0.06	0.14	0.14	0.09	0.09
Mean dependent variable	0.75	0.75	3.71	3.71	0.42	0.42

Regressions of self-reported newspaper readership on an indicator for the availability of a local Craigslist website in the county of the respondent, and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. Respondent controls include sex, age, an indicator for college degree and race. County controls include contemporaneous log population and number of Internet service providers, as well as share urban population, pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share - all measured in 2000 and interacted with year FEs. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

ing their subscription prices, which in turn would lead to lower demand. We test for this mechanism in Appendix figure C3 and table C4, looking at the impact of CL's entry on yearly subscription prices reported in the E&P yearbooks. With the exception of the last column, we obtain insignificant coefficients on both the indicator for CL-entry and its interaction with classified manager. There is hence no clear evidence of an increase in subscription prices.[26]

A second explanation is that readers respond to the changes in content brought about by CL's entry. One possibility is that the change toward less coverage of (local) politics that we document in Tables 5 and 6 alienated those readers interested in this type of content. Alternatively, perhaps the fall in circulation is driven by readers who were primarily interested in the classified ads which, after the entry of CL, became less appealing. Though in both cases some readers would ultimately be less exposed to news and political content, understanding which of these scenarios is more plausible can shed light on which segments of the population were most affected by the entry of CL.

To understand this question, it is useful to first get a sense how many readers reported reading different components of the newspaper prior to the entry of CL. To this end, we use data from a question included in the 1998-2001 waves of the GfK-MRI survey, which asked what sections of the newspaper the respondent would normally read (available for a sample of 100,519 individuals). Since each respondent could mention multiple sections and did not have to rank her choices, a clean comparison of the relative popularity of different sections is not possible. Still, the distribution of readers' preferences, depicted in Figure A9, is informative and indicates that most readers (63%) report reading the "General News" section (which includes politics), with the Sports and Business sections also being popular (38% and 37%, respectively). The Classified section is not far behind, however, with 34% of respondents reporting Classifieds as one of the sections they frequently read. It is, therefore, possible that a reduction in the value of print classifieds might be the driver of circulation declines.

To understand what types of readers drove the drop in readership of local papers following the entry of CL, we examine heterogeneity in the readership effect by propensity to read newspaper classified versus general news (including politics) sections. Using data from the 1998-2001 waves of the GfK-MRI survey, we estimate an elastic-net penalized regression model to identify the individual characteristics that are most predictive of reading the general

[26] Seamans and Zhu (2014) on the other hand find a significant increase in subscription prices in response to CL entry. Our analysis differs in using more complete data (i.e. covering all newspapers in the period 1995-2010) and a somewhat different empirical strategy, e.g. looking at reliance on classified ads at baseline rather than contemporaneously to account for its likely endogeneity.

news and the classifieds sections respectively.[27]

Based on the model estimated in the 1998-2001 data, we project two propensity scores for respondents in the later years. This procedure allows us to assign to each respondent in the post-CL surveys a probability for reading general news and one for reading classifieds. Projecting based on pre-CL data allows us to focus attention on differential changes among demographic types who would have been likely to read either classifieds or political news prior to CL entry, without the confound of the post-CL changes to newspapers' product. The projected propensity scores have fairly strong negative correlation (with correlation coefficient of -0.4), indicating that the groups that tend to read each section are relatively distinct.[28]

We then re-estimate the individual regression for the readership variable separately for two groups of respondents: i) those with above-median probability of reading classifieds and below-median probability of reading general news, and ii) those with below-median probability of reading classifieds and above-median probability of reading general news. The results, reported in Table 10, indicate that the decline in readership after the entry of CL is entirely driven by individuals with high news propensity and low classified propensity, and is again more pronounced where newspapers relied more heavily on classifieds ex ante. The effect is not significant for the high-classified-interest, low-news-interest group in any of the specifications.[29] Though newspapers which offered the most classifieds were most affected by the Craigslist shock, the readers least interested in classifieds reduced their newspaper reading most.

Taken together, these results support the view that the main driver of circulation reduction was the indirect changes in news content induced by newspapers' revenue loss following the entry of CL, rather than the direct effect of obsolescence and disappearance of print classified ads.

5.7. Political Outcomes

In the previous sections we have documented how newspapers affected by the entry of CL reported less about politics, in general, and local politics, in particular, and became less able to inform citizens about the misconduct of local officials. We also found that individuals in

[27] The characteristics most strongly associated with general news reading are white race, having a post-graduate degree, income in the 75K-150K range, being retired, being married 25 years or more, and being age 45-49. The characteristics most associated with classifieds reading are being unemployed, living in a small to moderate sized county, having a high school diploma only or "some college", being engaged (to be married), and being 25-29 years old.

[28] In particular, younger, less educated, less wealthy, and unemployed respondents report high interest in classifieds, whereas older, wealthier, more educated, and retired respondents report high interest in general news.

[29] Alternatively, in appendix table **??** we explore the two dimensions of readership propensity separately. We find more pronounced declines in readership for respondents with high rather than low news propensity and for respondents with low rather than high classified propensity.

TABLE 10: SELF-REPORTED NEWSPAPER READERSHIP: HETEROGENEITY

	(1)	(2)	(3)	(4)
	Low news propensity		High news propensity	
	High news propensity		Low classif. propensity	
Post-CL	-0.005	-0.004	-0.016**	0.009
	(0.007)	(0.012)	(0.008)	(0.012)
Post-CL × Classified Mgr.		0.001		-0.033**
		(0.013)		(0.013)
Respondent controls	Yes	Yes	Yes	Yes
County controls	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
Observations	84341	83475	84338	83854
Number of counties				
R^2	0.05	0.05	0.10	0.10
Mean dependent variable	0.32	0.32	0.54	0.54

Regressions of self-reported newspaper readership on an indicator for the availability of a local Craigslist website in the county of the respondent, and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. Respondent controls include sex, age, an indicator for college degree and race. County controls include contemporaneous log population and number of Internet service providers, as well as share urban population, pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share - all measured in 2000 and interacted with year FEs. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

areas affected by CL experienced a decline in readership of local papers not compensated by access to other news sources. Next we examine how these changes ultimately affected the behavior of voters and politicians. Indeed, given previous evidence on the relationship between press coverage, citizen electoral participation, and political accountability (Snyder and Strömberg, 2010), it is plausible to think that changes in coverage and readership may have ramifications for downstream political outcomes. We are specifically going to look at three outcomes that the literature on political accountability has traditionally focused on. For voters we look at voter turnout (Gentzkow et al., 2011; Gentzkow, 2006), and political polarization measured by split-ticket voting (Darr et al., 2018; Moskowitz, 2021; Trussler, 2020b). For politicians, we look at voting behavior of representatives in Congress (Snyder and Strömberg, 2010), with particular regard to ideological extremism and party discipline.

5.7.1. Turnout and split-ticket voting

As discussed in section 3, both electoral turnout and split-ticket voting are measured at the county level. Turnout is available for both presidential and midterm election years, while split vote only for presidential years since it captures precisely the probability of voting for candidates of different parties for president and for House representative. We assume that newspapers affect voters' behavior in the county in which they are based, which limits the sample to the 1,234 counties in which, according to the E&P data, at least one newspaper is headquartered. Since both outcomes are defined at county-level, we use estimating equation 4 as baseline specification. Given the considerable differences in the number of residents across counties, we weight observations by the county's total voting-age population.

In Table 11 we examine the effect of CL on electoral turnout. The results indicate that the entry of CL has a negative impact on turnout in presidential elections (columns 1-3), which is concentrated in counties where newspapers relied more heavily on classified ads. The coefficient on *Post-CL* $\times$ *Classif.Mgr* becomes smaller but remains significant even when controlling for $State \times Year$ fixed effects, which captures all time-varying factors common to counties in a given state, including the identity of presidential and senatorial candidates. In terms of magnitude, the effect corresponds to a drop between 1% and 2.5% of the sample mean, depending on the specification. We do not find similar effects for turnout in Midterm elections, which, traditionally, is much lower than in presidential years.

In Table 12 we examine the impact on split-ticket voting. The results indicate that following the entry of CL, voters became significantly less likely to split their votes between candidates of different parties. The effect is especially strong in areas where newspapers had more to lose from CL's competition, where the drop amounts to 16% of the sample mean. One interpretation of this finding is that, as less information about local representatives is available

TABLE 11: TURNOUT IN PRESIDENTIAL AND MIDTERM HOUSE ELECTIONS

	Presidential Turnout			Midterm House Turnout		
	(1)	(2)	(3)	(4)	(5)	(6)
Post-CL	0.004	0.008**	0.003	-0.014**	-0.010*	-0.006
	(0.004)	(0.003)	(0.002)	(0.006)	(0.006)	(0.013)
Post-CL × Classif. Mgr.	-0.013**	-0.011***	-0.005*	-0.0003	0.003	0.004
	(0.006)	(0.004)	(0.003)	(0.007)	(0.006)	(0.012)
County controls # 1	Yes	Yes	Yes	Yes	Yes	Yes
County controls # 2	Yes	Yes	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	No	Yes	Yes	No
State×Year FEs	No	No	Yes	No	No	Yes
Observations	4,903	4,903	4,903	4,832	4,832	4,832
R^2	0.95	0.96	0.98	0.86	0.88	0.73
Mean dependent variable	0.54	0.54	0.54	0.37	0.37	0.37

Regressions of turnout by county and election year on an indicator for the availability of a local Craigslist website, and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. OLS regressions in all columns. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. Observations are weighted by voting-age population. OLS regressions in all columns. Standard errors clustered by CL-area. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

from local media, voters rely more heavily on partisan cues - shaped primarily by the national political debate - when deciding on down-ballot races. This would lead to higher party alignment between different races, i.e. the "nationalization" of local elections (Moskowitz, 2021; Trussler, 2020b), and could further incite ideological divisions. This result connects to similar findings by Darr et al. (2018) regarding the impact of newspaper closures. We document that closures are not necessary to produce this effect: the impoverishment of local papers, and associated cuts in news coverage of local politics, are sufficient.

5.7.2. POLITICIANS' BEHAVIOR

Finally, in Table 13 we analyze how changes in news coverage and voters' behavior ultimately affected the conduct of politicians. In light of the significant decline in news coverage of congressional politics documented in section 5.4, we focus on Congress representatives. We look at the effect on ideological extremism, party discipline, and effort. As explained in section 4, for all outcomes pertaining to representatives, we collapse the variable *Post-CL* at the electoral district level, and interact it with the variable *Reader Share* which captures the overlap between the electoral district and the area served by local newspapers (Snyder and Strömberg, 2010).

TABLE 12: SPLIT-TICKET VOTING IN HOUSE VS. PRESIDENTIAL ELECTIONS

	Dependent variable: House-Pres R Vote Differential					
	(1)	(2)	(3)	(4)	(5)	(6)
Post-CL	-0.009	-0.013**	-0.013*	0.005	-0.0004	0.002
	(0.007)	(0.007)	(0.006)	(0.008)	(0.008)	(0.008)
Post-CL × Classif. Mgr.				-0.019*	-0.018*	-0.019**
				(0.011)	(0.009)	(0.009)
County controls # 1	Yes	Yes	Yes	Yes	Yes	Yes
County controls # 2	Yes	Yes	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	No	Yes	Yes	No
State×Year FEs	No	No	Yes	No	No	Yes
Observations	4,849	4,849	4,849	4,849	4,849	4,849
R^2	0.48	0.50	0.62	0.48	0.50	0.62
Mean dependent variable	0.12	0.12	0.12	0.12	0.12	0.12

Regressions of split-ticket voting by county and election year on an indicator for the availability of a local Craigslist website, and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. Split-ticket voting is defined as the absolute value of the difference between the Republican candidate vote share in the Presidential election and the Republican candidate vote share in the House election(s). County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Observations are weighted by voting-age population. OLS regressions in all columns. Standard errors clustered by CL-area. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In the first three columns of of Table 13 we look at the effect on ideological extremism. We find no clear evidence that representatives from areas served by CL became more ideologically extreme. Though the coefficients of interest are positive, they are small and not statistically significant in all specifications.

Next we look at party discipline, measured by the share of roll-call votes on which a representative voted with the leadership of his/her party. Our results, reported in columns 4-6 of Table 13, indicate that, after the entry of CL, local representatives displayed significantly higher levels of party discipline, especially in areas with high-reader share where changes in newspaper content are more likely to learn about the behavior of their elected officials. The coefficient on the interaction $Post-CL\times$ *Reader Share* remains significant at the 1% level even when controlling for $State \times Year$ fixed effects. The effect is also quite sizeable: looking at the most complete specification (column 6), in places where newspaper market and electoral district coincide, the entry of CL is associated with an increase in party discipline of about 1.5% of the sample mean which is 83%.

Finally, we study the effect of CL on the number of bills sponsored by representatives in a

year, a proxy for the effort they exert in their legislative activity beyond roll-call voting. The results, presented in columns 7-9 of Table 13, show some evidence of a negative impact of CL on effort in areas with high-reader share, which is statistically significant when controlling for *State timesYear* fixed effects, that is when comparing representatives elected in different districts in the same state, a rather compelling comparison. The effect is quite large: where newspaper market and electoral district coincide, the entry of CL is associated with a drop in the number of sponsored bills equal to 3.3% of the sample mean.

Taken together, our findings point that the profound transformations triggered by the entry of CL had tangible effects on the behavior of politicians favoring the widening of the ideological divide between parties at the expense of electoral accountability.

TABLE 13: BEHAVIOR OF REPRESENTATIVES

	ADA Extremity			Vote w/ Party Leadership			No. Bills Sponsored		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post-CL	0.003	0.002	0.0007	0.012***	-0.0003	-0.001	-0.045	0.034	0.115
	(0.002)	(0.002)	(0.002)	(0.005)	(0.004)	(0.003)	(0.119)	(0.114)	(0.103)
Post-CL × Reader Share	0.004	0.003	0.001	0.027***	0.011**	0.013***	-0.252	-0.142	-0.309*
	(0.003)	(0.003)	(0.002)	(0.005)	(0.005)	(0.004)	(0.170)	(0.162)	(0.161)
County controls # 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls # 2	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
State×Year FEs	No	No	Yes	No	No	Yes	No	No	Yes
Observations	41,022	41,022	41,022	41,510	41,510	41,510	41,594	41,594	41,594
R^2	0.44	0.46	0.55	0.64	0.73	0.78	0.48	0.50	0.56
Mean dependent variable	0.88	0.88	0.88	0.83	0.83	0.83	9.2	9.2	9.2

OLS regressions in all columns. Standard errors clustered by CL-area. An observation is a district-newspaper-year, and observations are weighted by the newspaper's market share in the district. Market shares sum to one, such that each district-year gets total weight of one. "Reader Share" is the fraction of the paper's subscribers who reside in the district. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6. CONCLUSION

Hamilton (2004) lays out the basic economics of the news-gathering business: high fixed costs - in the form of reporting staff who must develop expertise in their subjects and form long-term relationships with their sources - combined with a non-excludable product lead generally to under-provision of news production relative to the social optimum. Counteracting this unhappy equilibrium to some degree are reporters' professional norms, which value the production of "hard news" and investigative journalism over cheaper-to-produce

and sometimes more popular "soft news."

For a time in the 20th century, local monopoly papers were able to extract sizable profits from the advertising business. Reporters employed by those papers captured some of these rents in the form of resources dedicated to reporting of local political news and other "hard" topics valued by journalists themselves (rather than readers or advertisers). The growth of advertising profits, in fact, can be directly tied to the emergence of the ideal of an independent press staffed by professional journalists, in contrast to the 19th-century norm of newspapers operated as propaganda organs of local party organizations (Petrova, 2011).

The emergence of competition in the advertising business from new internet-based entrants in the early 2000s upset this tenuous balance, eliminating the economic profits which had supported investments in money-losing but high-prestige reporting. We show that the entry of one particularly important such competitor, the classified advertising platform Craigslist, had severe impacts on newspapers' staffing levels and production of news coverage relating to local politics.

The Craigslist effect is not simply a consequence of changes to the demand for news induced by internet availability; rather, it appears to operate by reducing newspapers' ability to invest in local reporting resources. Papers that were especially reliant on classified advertising in the pre-Craigslist period saw much larger changes on these dimensions than comparably internet-exposed but less classified-dependent papers. The loss of advertising revenues at these papers seems to have particularly reduced coverage of local representatives and investigative coverage, two areas with large positive externalities but also large private costs for newspaper operators.

Consistent with existing work on media effects on political outcomes, we find that there were measurable social consequences of this change in the production of news content. Voters in areas served by papers affected most by the Craigslist shock saw their House elections become more nationalized, which we interpret as a consequence of thinner information about the local incumbent's behavior. Changes in the media landscape may thus be an important driver of the overall trend towards nationalization of elections in the United States (Hopkins, 2018).

The change in voters' information about their representative's performance in office had consequences for elected officials' behavior as well. We show that the reduction of representative-specific information led House members' votes to more closely track the priorities of their (polarized) national leadership. These results are consistent with other work on the relationship of legislators' behavior to the quality of voters' information (Snyder and Strömberg, 2010; Trussler, 2020a).

Our results have implications for our understanding of the link between advertising market

structure and the market for news. They highlight the fragility of compensating the production of a public good - politically relevant information - with proceeds from bundled advertising. Technological innovation that unbundles the two products, as Craigslist did for classified advertising, can have spillover effects on the news market, with significant and lasting consequences for the quality of representation and political polarization.

References

Adler, E. S. and J. Wilkerson (2020): "Congressional Bills Project: 1973-2014," .

Anderson, S. and P. Habel (2009): "Revisiting Adjusted ADA Scores for the U.S. Congress, 1947–2007," *Political Analysis*, 17, 83–88.

Angelucci, C., J. Cage, , and M. Sinkinson (2020): "Media Competition and News Diets," *NBER Working Paper No.26782*.

Besley, T. and R. Burgess (2002): "The Political Economy of Government Responsiveness: Evidence from India," *Quarterly Journal of Economics*, 117, 1415–1451.

Bhuller, M., T. Havnes, J. McCauley, and M. Mogstad (2020): "How the Internet Changed the Market for Print Media," .

Campante, F., R. Durante, and F. Sobbrio (2017): "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation," *Journal of the European Economic Association*, 16, 1094–1136.

Canes-Wrone, B., D. W. Brady, and J. F. Cogan (2002): "Out of step, out of office: Electoral accountability and House members' voting," *American Political Science Review*, 96, 127–140.

Chan, J. and A. Ghose (2014): "Internet's dirty secret: assessing the impact of online intermediaries on HIV transmission," *MIS Quarterly*, 38.

Cunningham, S., G. DeAngelo, and J. Tripp (2019): "Craigslist's Effect on Violence Against Women," .

Darr, J. P., M. P. Hitt, and J. L. Dunaway (2018): "Newspaper closures polarize voting behavior," *Journal of Communication*, 68, 1007–1028, publisher: Oxford University Press.

de Chaisemartin, C. and X. D'Haultfoeuille (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, forthcoming.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012): "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102, 2955–80.

Falck, O., R. Gold, and S. Heblich (2014): "E-lections: Voting Behavior and the Internet," *American Economic Review*, 117, 2238–2265.

FCC (2016): "Federal Communications Commission Report FCC 16-107," Tech. rep.

Gallagher, R. J., K. Reing, D. Kale, and G. V. Steeg (2017a): "Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge," *Transactions of the Association for Computational Linguistics*, 5, 529–542.

——— (2017b): "Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge," *Transactions of the Association for Computational Linguistics*, 5, 529–542.

GAO, P., C. LEE, AND D. MURPHY (2020): "Financing Dies in Darkness? The Impact of Newspaper Closures on Public Finance," *Journal of Financial Economics*, 135, 445–467.

GAVAZZA, A., M. NARDOTTO, AND T. VALLETTI (2019): "Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies," *Review of Economic Studies*, 86, 2092–2135.

GENTZKOW, M. (2006): "Television and voter turnout," *Quarterly Journal of Economics*, 121.

GENTZKOW, M. AND J. M. SHAPIRO (2010): "What Drives Media Slant? Evidence from U.S. Daily Newsapers," *Econometrica*, 78, 35–71.

GENTZKOW, M., J. M. SHAPIRO, AND M. SINKINSON (2011): "The effect of newspaper entry and exit on electoral politics," *American Economic Review*, 101(7).

HAMILTON, J. (2004): *All the news that's fit to sell: How the market transforms information into news*, Princeton University Press.

HAMILTON, J. T. (2016): *Democracy's Detectives: The Economics of Investigative Journalism*, Cambridge, Mass.: Harvard University Press.

HOPKINS, D. J. (2018): *The increasingly United States: How and why American political behavior nationalized*, University of Chicago Press.

KROFT, K. AND D. G. POPE (2014): "Does online search crowd out traditional search? Evidence from Craigslist," *Journal Labor Economics*, 32(2), 259–303.

LARCINESE, V. AND L. MINER (2018): "Was Obama Elected by the Internet? Broadband Diffusion and Voters' Behavior in US Presidential Elections," .

LEWIS, J., K. POOLE, H. ROSENTHAL, A. BOCHE, A. RUDKIN, AND L. SONNET (2020): "Voteview: Congressional roll-call votes database," .

MAHONE, J., Q. WANG, P. NAPOLI, M. WEBER, AND K. MCCULLOUGH (2019): "Who's Producing Local Journalism? Assessing Journalistic Output Across Different Outlet Types," Tech. rep.

MANACORDA, M. AND A. TESEI (2020): "Liberation Technology: Mobile Phones and Political Mobilization in Africa," *Econometrica*, 88, 533–567.

MAYHEW, D. R. (1974): *Congress: The Electoral Connection*, Yale University Press.

MCCHESNEY, R. W. AND J. NICHOLS (2011): *The death and life of American journalism: The media revolution that will begin the world again*, Bold Type Books.

MOSKOWITZ, D. J. (2021): "Local news, information, and the nationalization of US Elections," *American Political Science Review*, 115, 114–129.

PETERSON, E. (2020): "Paper Cuts: How Reporting Resources Affect Political News Coverage," .

PETROVA, M. (2011): "Newspapers and parties: How advertising revenues created an independent press," *American Political Science Review*, 790–808.

SEAMANS, R. AND F. ZHU (2014): "Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers," *Management Science*, 60(2), 265–540.

SNYDER, J. M. AND D. STRÖMBERG (2010): "Press coverage and political accountability," *Journal of Political Economy*, 118(2), 355–408.

STARKMAN, D. (2014): *The watchdog that didn't bark: The financial crisis and the disappearance of investigative journalism*, Columbia University Press.

TRUSSLER, M. (2020a): "The Effects of High-Information Environments on Legislative Behavior in the US House of Representatives," *Legislative Studies Quarterly*.

——— (2020b): "Get Information or Get in Formation: The Effects of High-Information Environments on Legislative Elections," *British Journal of Political Science*.

A. Background and Data

Figure A1: Evolution of newspaper revenues by source

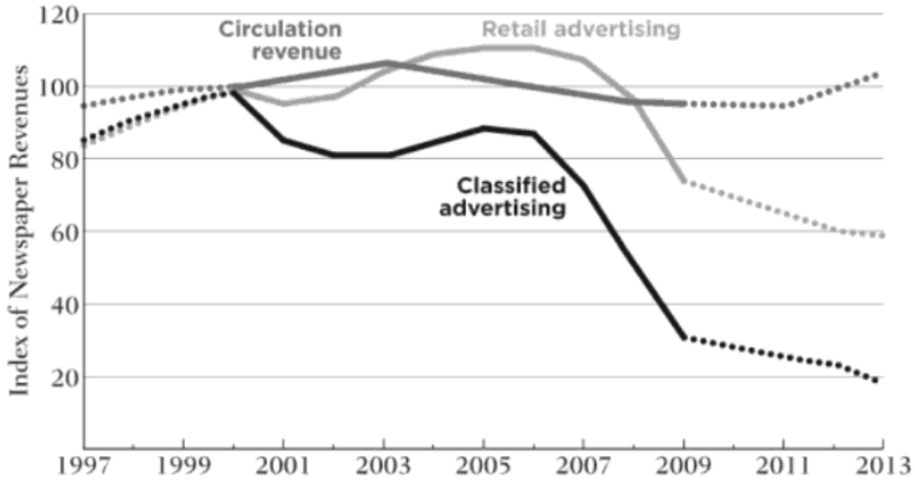


Figure 4: U.S. Newspaper Revenues over Time (Index: year 2000 = 100).

Index of newspaper revenues from circulation, retail advertising and classified advertising - 1997 to 2013. Source: Newspaper Association of America.

FIGURE A2: CRAIGSLIST: LAYOUT IN 2000 AND 2016

2000

craigslist

san francisco bay area

other craigslists ▾ go

help? post a listing
FAQ subscriptions

search craigslist

community ▾
search

feedback

our policies

about craigslist

questions@craigslist.org

nonprofit venture forum

updated 19 June

community & events
events / entertainment
tech events
classes / workshops
artists / musicians
community
pets / animals
volunteers

personals
women for women
women for men
men for women
men for men
misc romance

activity partners
carpool / rideshare

discussion boards

housing
apts / housing
apts / housing wanted
rooms / shared
rooms / shared wanted
sublets / temporary / vac
office / commercial
parking / storage

sale / wanted
barter / swap / free
bikes / cycles / scooters
cars / trucks
computer / tech stuff
general for sale
items wanted
small biz ads
tickets

resumes
freelance services 1099

jobs
accounting / finance
admin / customer service
architect / engineer / CAD
arts / print / design
business / e-biz / mgmt
human resources
internet / web engineering
legal / paralegal
marketing / advertising / pr
medical / health / biotech
network / telecomm / WAN
nonprofit sector
retail / hospitality / food
sales / biz dev
software / QA / DBA / etc
system administration
technical support
tv / film / video / radio
web / info design
writing / editing
et cetera

2016

craigslist

post to classifieds
my account

search craigslist
search

event calendar

M	T	W	T	F	S	S
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	1	2	3	4
5	6	7	8	9	10	11

help, faq, abuse, legal
avoid scams & fraud
personal safety tips
terms of use
privacy policy
system status

about craigslist
craigslist is hiring in sf
craigslist open source
craigslist blog
best-of-craigslist
craigslist TV
"craigslist joe"
craig connects
progressive directory

weather quake tide

SF bay area w sfc sby eby pen nby scz

community
activities, artists, childcare, classes, events, general, groups, local news, lost+found, musicians, pets, politics, rideshare, volunteers

personals
strictly platonic
women seek women
women seeking men
men seeking women
men seeking men
misc romance
casual encounters
missed connections
rants and raves

discussion forums
apple, arts, atheist, autos, beauty, bikes, celebs, comp, crafts, diet, divorce, dying, eco, educ, feedbk, film, fitness, fixit, food, frugal, gaming, garden, haiku, help, history, housing, jobs, jokes, kink, legal, linux, m4m, manners, marriage, media, money, motocy, music, nonprofit, open, outdoor, over 50, parent, pets, philos, photo, p.o.c., politics, psych, queer, recover, religion, romance, science, spirit, sports, tax, travel, tv, vegan, w4w, wed, wine, women, words, writing, yoga

housing
apts / housing
housing swap
housing wanted
office / commercial
parking / storage
real estate for sale
rooms / shared
rooms wanted
sublets / temporary
vacation rentals

for sale
antiques, appliances, arts+crafts, atv/utv/sno, auto parts, baby+kid, barter, beauty+hlth, bikes, boats, books, business, cars+trucks, cds/dvd/vhs, cell phones, clothes+acc, collectibles, computers, electronics, farm+garden, free, furniture, garage sale, general, heavy equip, household, jewelry, materials, motorcycles, music instr, photo+video, rvs+camp, sporting, tickets, tools, toys+games, trailers, video gaming, wanted

services
automotive, beauty, cell/mobile, computer, creative, cycle, event, farm+garden, financial, household, labor/move, legal, lessons, marine, pet, real estate, skilled trade, sm biz ads, therapeutic, travel/vac, write/ed/tran

jobs
accounting+finance
admin / office
arch / engineering
art / media / design
biotech / science
business / mgmt
customer service
education
food / bev / hosp
general labor
government
human resources
internet engineers
legal / paralegal
manufacturing
marketing / pr / ad
medical / health
nonprofit sector
real estate
retail / wholesale
sales / biz dev
salon / spa / fitness
security
skilled trade / craft
software / qa / dba
systems / network
technical support
transport
tv / film / video
web / info design
writing / editing
[ETC]
[part-time]

gigs
computer, creative, crew, domestic, event, labor, talent, writing

resumes

english

nearby cl
bakersfield, chico, fresno, gold country, hanford, humboldt, inland empire, klamath falls, las vegas, los angeles, medford, mendocino co, merced, modesto, monterey, orange co, palm springs, redding, reno, roseburg, sacramento, san luis obispo, santa barbara, santa maria, siskiyou co, stockton, susanville, ventura, visalia-tulare, yuba-sutter

us cities
us states
canada
cl worldwide

FIGURE A3: CRAIGSLIST: LAYOUT OF HOUSING LISTINGS

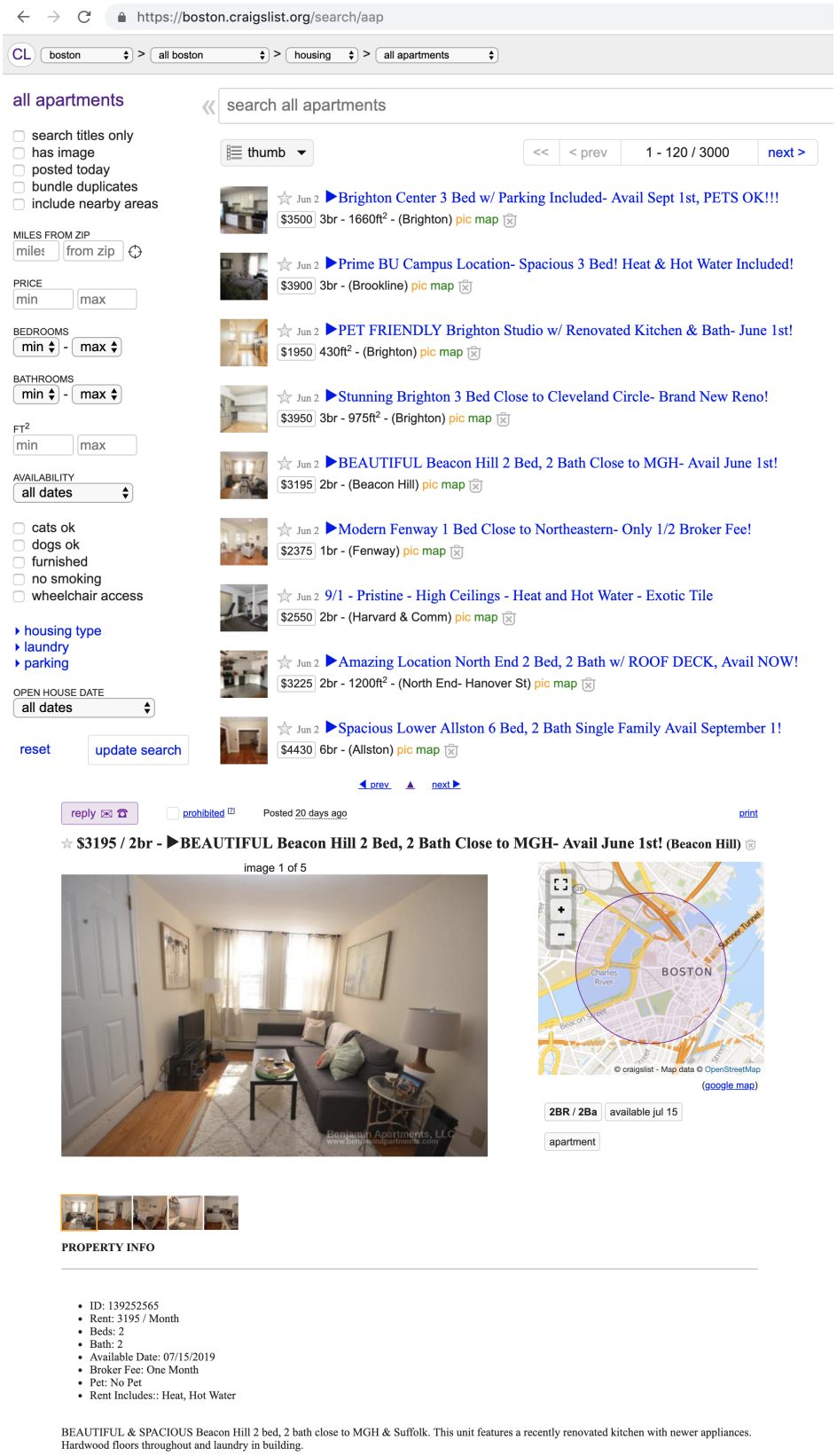
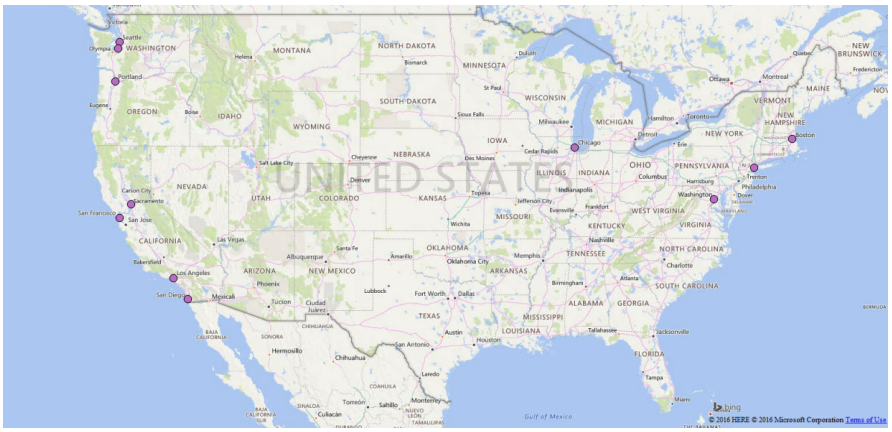
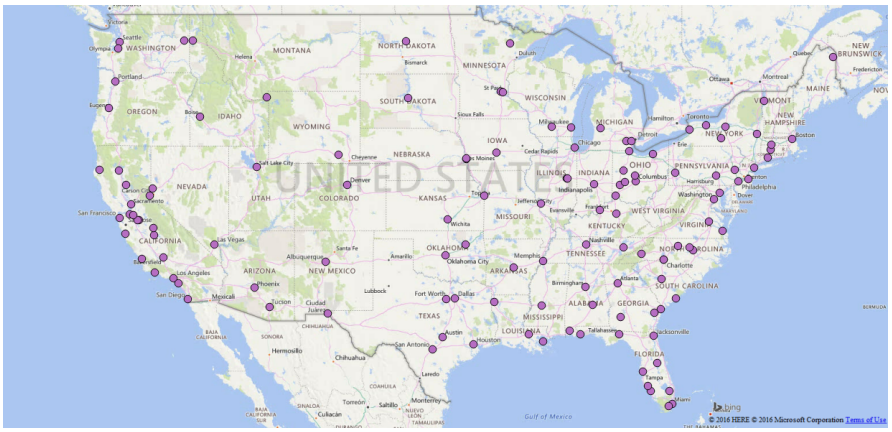


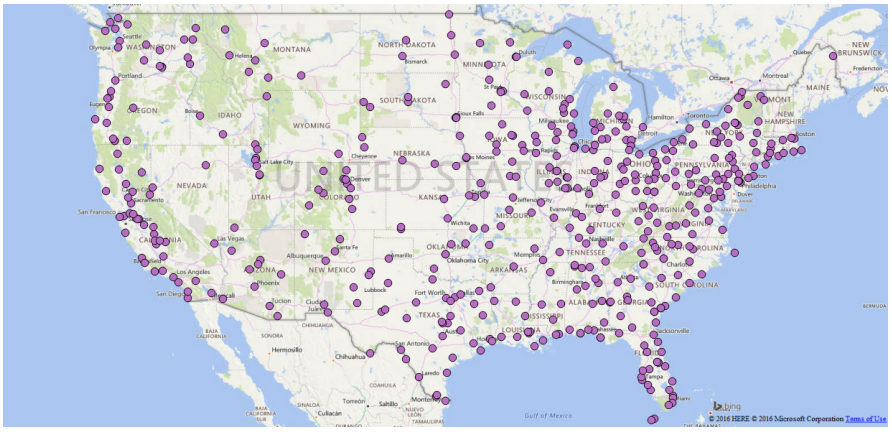
FIGURE A4: GEOGRAPHIC DISTRIBUTION OF CL WEBSITES (2000-2010)



(a) 2000



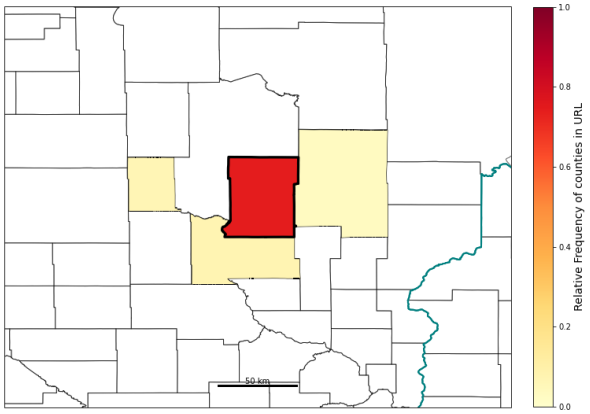
(b) 2005



(c) 2010

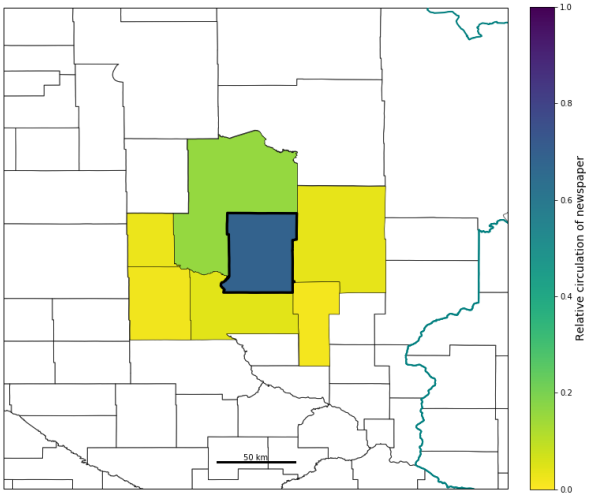
Distribution of local CL websites in the years 2000, 2005 and 2010.

FIGURE A5: DISTRIBUTION OF ADS POSTED ON HTTPS://BRAINERD.CRAIGSLIST.ORG/



Geographic distribution of the location of ads posted in the housing, jobs and sales sections of https://brainerd.craigslist.org/ in years 1 and 2 after the website opening. Source: Internet Archive.

FIGURE A6: DISTRIBUTION OF CIRCULATION OF THE BRAINERD DISPATCH



Geographic distribution of the circulation of the *Brainerd Dispatch* in 2002. Source: Alliance for Audited Media.

FIGURE A7: EXTRACT FROM THE EDITOR AND PUBLISHER YEARBOOKS

The Reporter
(m-mon to fri; m-sat)
The Reporter, 307 Derstine Ave.; PO Box 390, Lansdale, PA 19446; gen tel (215) 855-8440; adv tel (215) 361-8849; ed tel (215) 361-8814; gen fax (215) 855-6147; ed fax (215) 855-3432; adv email imaging@thereporteronline.com; ed email letters@thereporteronline.com; web site http://www.thereporteronline.com.
Group: Journal Register Co.
Circulation: 17,808(m); 15,590(m-sat); ABC Sept. 30, 2003.
Price: $0.50(d); $0.50(sat); $3.00/wk (carrier); $156.00/yr (carrier), $196.00/yr (mail).
Advertising: Open inch rate $33.83(m); $33.83(m-sat). **Representatives:** Landon Media Group; U.S. Suburban Press Inc.; Robert Hitchings & Co.
News Services: AP, GNS.
Politics: Independent. **Established:** 1870.

CORP. MGMT./GEN. MGMT.

Pres./Pub.	Al Frattura
Controller/Purchasing Agent	Bernard DeAngelis

ADVERTISING SALES MGMT.

Adv. Dir.	Robert Twesten
Display Adv. Mgr.	Angel Hernandez

NEWS EXECUTIVES

Exec. Ed.	Nona Breaux

EDITORIAL MGMT.

City Ed.	Monica Thompson
Lifestyles Ed.	Aixa Torregrosa
Night Ed.	Linda Doell
Page 1 Ed.	Dan Sharer
Chief Photographer	Geoff Patton
Special Sections	Kass Picozzi
Sports Ed.	Kevin Lilley

The Reporter, Landsdale PA

Chicago Tribune
(m-mon to tues; m-wed to fri; m-sat; S)
Chicago Tribune, 435 N. Michigan Ave., Chicago, IL 60611; gen tel (312) 222-3232; gen fax (312) 222-2595; gen email tribletter@tribune.com; web site http://www.chicagotribune.com.
Group: Tribune Co.
Circulation: 680,879(m); 512,455(m-mon to tues); 571,576(m-sat); 1,002,166(S); ABC Sept. 30, 2003.
Price: $0.50(d); $0.50(sat); $1.79(S); $4.40/wk; $228.80/yr.
Advertising: Open inch rate $580.00(m); $580.00(m-sat); $842.00(S). **Representatives:** Western States Associates Inc.
News Services: AP, RN, NYT, TMS, DJ, KRT.
Politics: Independent. **Established:** 1847.
Advertising not accepted: Handguns, ammunition and tobacco.

CORP. MGMT./GEN. MGMT.

Pres./Pub./CEO	Scott C. Smith
Sr. Vice Pres./Gen. Mgr.	Richard Malone
Sr. Vice Pres./Ed.	Ann Marie Lipinski
Vice Pres., Circ./Consumer Mktg.	Vincent Casanova
Vice Pres./Chief Tech. Officer	Darko Dejanovic
Vice Pres., Adv. Mktg./Sales	Ken DePaola
Vice Pres., Finance	Phil Doherty
Vice Pres., Human Resources	Janice Jacobs
Vice Pres., Devel.	Owen Youngman
Vice Pres./Dir., Opns.	Tony Hunter
Gen. Mgr., Chicago Tribune Interactive	Alison Scholly
Dir., Technical Devel.	Scott Tafelski
Dir., Technical Opns./Help Desk	Robert Trinchet
Dir., Client Servs.	Deepak Agarwal

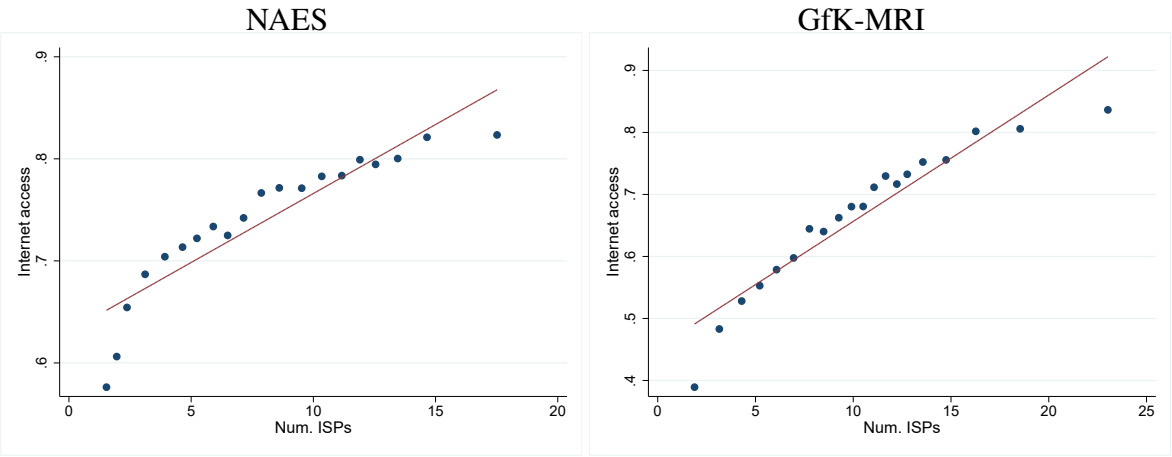
ADVERTISING SALES MGMT.

Dir., Nat'l Adv.	Dan Dunn
Dir., Network Adv.	Ron Goldberg
Dir., Classified Adv.	Barbara Swanson
Dir., Major Accts.	Douglas Thomas

'phics[width=0.45]"background/chicago$_{t}$*ribune*3$_{4}$"

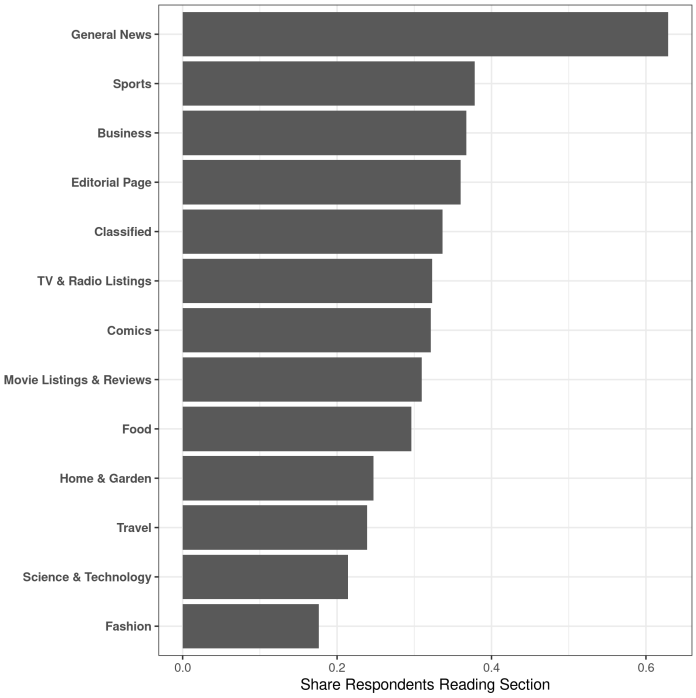
The Chicago Tribune

FIGURE A8: NUMBER OF ISPS AS A PROXY FOR LOCAL INTERNET PENETRATION



Binned scatter plot for the relationship between number of ISPs available in the county of the respondent and self-reported Internet access at home (GfK-MRI) and at home or at work (NAES).

FIGURE A9: NEWSPAPER SECTIONS BY READERSHIP (1998-2001)



Source: GfK-MRI Survey

B. Additional Data Details

This section contains additional details and summary statistics on the data sets used in the analyses, as well as measurement validation exercises.

B.1. Validating the classified manager proxy

We validate the classified manager indicator as a proxy for classified intensity using data from the website *Newspapers.com*, which archives digitized historical copies of newspapers. We located 261 papers in our dataset which appear in the Newspapers.com archive. For each of these papers, we sampled the edition of the paper published on the first Sunday of each month in all years from 1995 until the year prior to CL entry, substituting another day when the Sunday edition was not available.

We measure classified intensity as the number of pages on which the term "Classified" appears, divided by the total number of pages in the issue. We collected this measure for a total of 43,400 issues across the 261 papers available in the Newspapers.com archive. Prior to the entry of Craigslist in a market, the median issue in our sample had 8 pages of classified advertising, or 15% of the issue's total page count.

Table B1 shows the results of regressions where the outcome is the fraction of pages in the newspaper's print edition that contain classified advertising. The left two columns use all days with weekday fixed effects to control for cyclical variation in classified intensity, while the last two use Sundays only (traditionally the biggest day for classified advertising).

Table B1: Share of Pages Devoted to Classified Ads in Pre-CL Period, by Existence of Classified Manager in 2000.

	Classified Page Share		Classified Page Share (Sun)	
	(1)	(2)	(3)	(4)
Classif. Mgr.	0.031*	0.022	0.037**	0.028
	(0.016)	(0.015)	(0.018)	(0.017)
Pop / Circ controls	No	Yes	No	Yes
Year FEs	Yes	Yes	Yes	Yes
Day-of-Week FEs	Yes	Yes	No	No
Observations	30,112	29,952	23,280	23,154
R^2	0.01	0.03	0.01	0.02
Mean dependent variable	0.20	0.20	0.21	0.21

OLS regressions in all columns. Standard errors clustered by newspaper.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B.2. Details on Newspaper Content Processing

This section contains details on the procedures used to process raw text content from the newspapers in our sample to construct lower-dimensional representations of the content. Source data are from the *NewsBank* database. We conduct two main kinds of processing on text data: keyword searches and topic modeling. Keyword searches use the full database containing more than 100M full-text articles, while our topic model uses a smaller random sample of about 2M articles consisting of all articles published on 10 randomly sampled dates in each newspaper-year between 1999 and 2010. The topic-modeling sample limits to the first paragraph of text, plus the headline.

B.2.1. Keyword searches

Politician names Our first set of keyword searches look for the names of representatives in the US House. We use a list of representatives from https://github.com/unitedstates/congress-legislators, and for each member construct a (case-insensitive) regular expression of the form `"(congres.*|rep.*) FIRSTNAME LASTNAME"`. This expression matches strings like "Rep. Adam Smith" or "Congressman Adam Smith" but not "Adam Smith" alone. We require the inclusion of the title to cut down on false positives, as many members of Congress have common names. This does introduce the possibility of false negatives, but we have found that articles covering a member usually include the title and full name at first mention before switching to a shorter form like "Mr. Smith".

We count the number of *articles*[30] in which the pattern described above appears on each newspaper-day, and then aggregate to the level of congressman by newspaper by year.

Politician titles We next construct keyword searches for common titles of politicians. The complete list of terms is given in Table B2. We aggregate counts (again measured in article rather than individual mention terms) to the newspaper-year level.

Corruption keywords Finally, we conduct a set of keyword searches for terms related to coverage of public corruption cases. We measure this coverage in two ways. First, we search for the names of the most common charges brought in prosecutions of government officials in the Transactional Records Access Clearinghouse (TRAC) criminal enforcement database.[31] The keywords (in regular expression form) that we search for are: `conspir(acy|e|ator)`, `fraud`, `bribe?(ry|ing|d)?`, `racketeer`, `hobbs act`, `misappropriat`, `[ ]rico[ .,]`,

[30] I.e., each article that mentions the congressperson at all counts as 1, regardless of how many times the congressman is referenced in the article.

[31] https://tracfed.syr.edu/

TABLE B2: KEYWORDS FOR POLITICIAN TITLE SEARCHES

Regular Expression	Example Match
`mayor`	Mayor
`councilperson`	Councilperson
`council(wo)?man`	Councilman
`council member`	Council Member
`state (sen(ator|.)|rep(resentative|.))`	State Sen.
`governor`	Governor
`alderman`	Alderman
`commissioner`	Commissioner
`(city|county) manager`	City Manager

`wire fraud`, `mail fraud`, `color of law`, `launder`, `embezzl`, `perjur`, `extort`, and `kickback`. We sum the counts across all phrases and aggregate to the level of newspaper-year.

In addition to the terms from TRAC, we also use a set of words and phrases that Hamilton (2016) includes in a typology of principal-agent problems that investigative journalists seek to expose. These phrases are the basis of Hamilton's study of the correlates of investigative intensity across papers and over time. To match Hamilton, we use his exact keywords rather than regular expressions that could capture variants of the same phrases. There is some overlap with the TRAC terms, but Hamilton's definition of principal-agent problems also includes behaviors that are not necessarily federal crimes, such as nepotism or mismanagement. The phrase list[32] is: "wasteful," "mismanag," "neglect," "bribery," "embezzlement," "steal," "corruption," "nepotism," "patronage," "conflict of interest," "rent seeking," "influence peddling," "favoritism," "abuse," "harassment," "misconduct," "discrimination," "misuse," "fraud," "deception," and "mislead." We sum the counts across all phrases and aggregate to the level of newspaper-year.

B.3. TOPIC MODELS

Our method for extracting the topical coverage of affected newspapers follows Gallagher et al.'s 2017b Correlation Explanation (CorEx) method. This is a semi-supervised method that allows input of a minimal set of "anchor" words, and then finds topics by searching for groups of words that co-occur with the anchors. We apply this method to the text of a random sample of 2 million articles from the NewsBank corpus.

We use the semi-supervised method rather than the more traditional unsupervised Latent Dirichlet Allocation (LDA) because it allows us to focus on specific topics of interest. We are interested in separating various dimensions of political news coverage: coverage related

[32] Taken from Hamilton (2016), pp. 37.

FIGURE B1: TOP TEN WORDS ASSOCIATED WITH EACH COREX TOPIC. ANCHOR WORDS ARE UNDERLINED.

```
0: presid,feder,govern,compani,tax,washington,percent,increas,pai,billion
1: council,mayor,board,plan,student,educ,fund,commun,project,program
2: repres,senat,congress,republican,elect,democrat,vote,candid,polit,gov
3: intern,war,foreign,iraq,militari,movi,film,american,soldier,terrorist
4: man,kill,injuri,injur,accid,crash,woman,diseas,victim,suffer
5: music,art,food,festiv,featur,concert,event,artist,band,holidai
6: car,vehicl,driver,road,truck,traffic,highwai,drive,mile,street
7: di,born,funer,son,daughter,church,surviv,servic,cemeteri,obituari
8: game,team,coach,win,season,plai,victori,footbal,score,player
9: polic,charg,court,arrest,judg,investig,attornei,accus,sheriff,suspect
```

to local, congressional, national and foreign politics. We seed separate anchors for these 4 topics, and run the CorEx model with 10 topics in total. Figure B1 presents the resulting topics, as described by their most representative words, with our anchor-words highlighted in bold. The 10 resulting topics can be labeled as follows: local politics, congressional politics, national politics, foreign politics, entertainment, health / family, weather, crime, obituaries. For each of the 2 million articles in the corpus, the CorEx model outputs a set of 10 unconditional probabilities for the article belonging to that a topic. Importantly, these probabilities do not necessarily sum to 1 - an article can simultaneously belong to more than one topic, or to none. To examine the effects of CL's entry, we aggregate the distribution of probabilities by newspaper and year, and estimate the standard diff-in-diff equations specified in section 4, with the average probability for each one of the 10 topics as dependent variable.

C. Additional results

Table C1: Number of Jobs: Robustness to controlling for newspaper size

	Dependent variable: Newspaper number of jobs							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	-1.049*	-0.523	-0.753	-0.893	0.303	0.667	0.451	0.278
	(0.595)	(0.558)	(0.596)	(0.735)	(0.482)	(0.484)	(0.508)	(0.599)
Post-CL × Circ. 2000	-0.016	-0.011	-0.009	-0.007	-0.012	-0.007	-0.006	-0.004
	(0.013)	(0.014)	(0.014)	(0.015)	(0.013)	(0.014)	(0.015)	(0.016)
Post-CL × Classified Mgr.					-3.006***	-2.657***	-2.667***	-2.623***
					(0.668)	(0.594)	(0.613)	(0.698)
Log population, #ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	22516	22516	22515	21762	22499	22499	22498	21745
Number of newspapers	1506	1506	1506	1460	1504	1504	1504	1458
R^2	0.90	0.91	0.91	0.92	0.90	0.91	0.91	0.92
Mean dependent variable	21.38	21.38	21.38	21.22	21.39	21.39	21.39	21.24

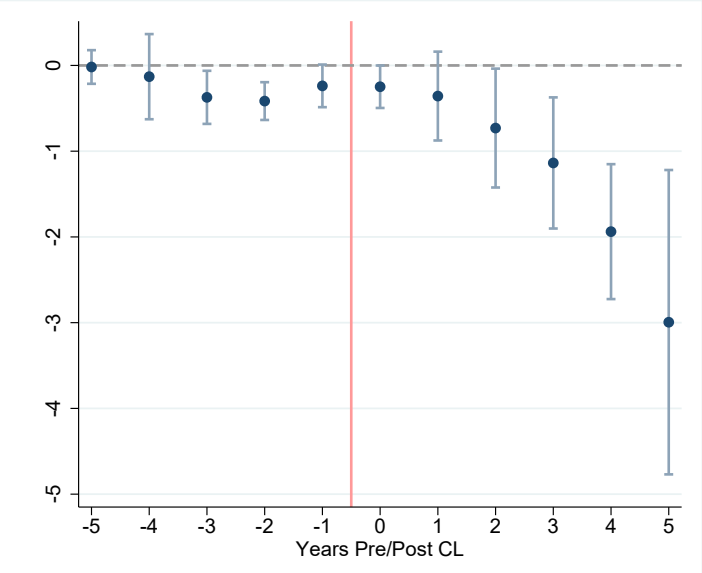
Regressions of number of jobs by newspaper and year on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C2: NUMBER OF STAFF

	Dependent variable: Newspaper number of employees							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	-1.701***	-1.012***	-1.107***	-1.086***	-0.112	0.260	0.131	0.070
	(0.339)	(0.329)	(0.344)	(0.338)	(0.341)	(0.339)	(0.345)	(0.349)
Post-CL × Classified Mgr.					-3.010***	-2.490***	-2.405***	-2.299***
					(0.513)	(0.480)	(0.477)	(0.516)
Log population, #ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	22797	22797	22796	22044	22583	22583	22582	21832
Number of newspapers	1540	1540	1540	1492	1505	1505	1505	1459
R^2	0.92	0.92	0.92	0.93	0.92	0.92	0.92	0.93
Mean dependent variable	17.84	17.84	17.84	17.68	17.90	17.90	17.90	17.74

Regressions of number of employees by newspaper and year on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE C1: NUMBER OF STAFF – EVENT STUDY



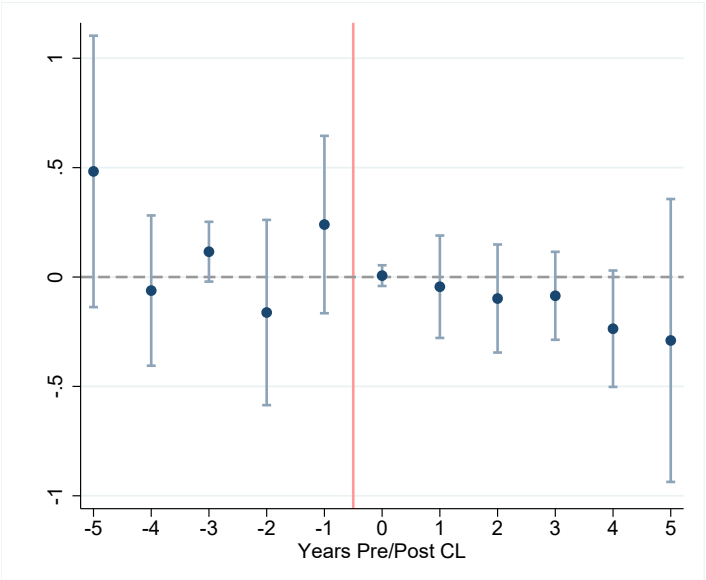
Dynamic effect of the availability of a local CL website on number of employees by newspaper and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

TABLE C3: NUMBER OF PAGES

	Dependent variable: Number of pages published							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	0.293	0.250*	0.125	0.259*	0.140	0.100	-0.026	0.136
	(0.191)	(0.139)	(0.171)	(0.151)	(0.128)	(0.126)	(0.162)	(0.213)
Post-CL × Classified Mgr.					0.286	0.292	0.289	0.238
					(0.316)	(0.286)	(0.294)	(0.347)
Log population, #ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	21637	21637	21637	20921	21509	21509	21509	20793
Number of newspapers	1471	1471	1471	1426	1450	1450	1450	1406
R^2	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.98
Mean dependent variable	28.50	28.50	28.50	28.31	28.52	28.52	28.52	28.34

Regressions of number of pages published by newspaper and year on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE C2: NUMBER OF PAGES – EVENT STUDY



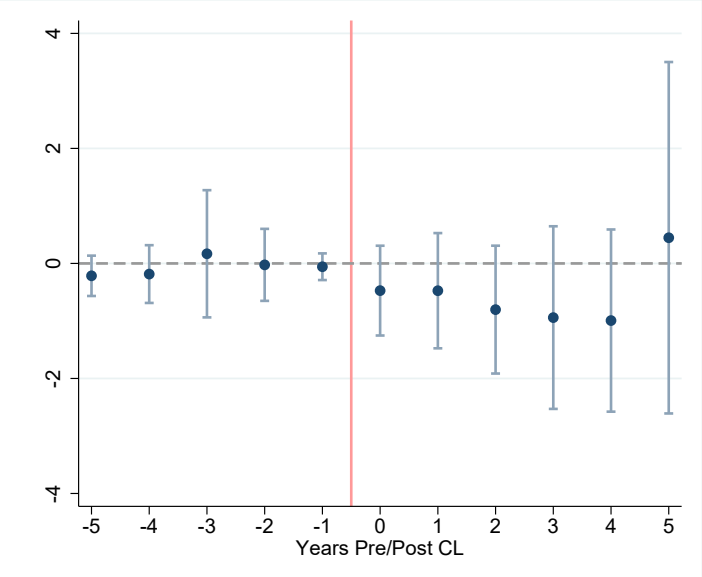
Dynamic effect of the availability of a local CL website on number of pages published by newspaper and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

TABLE C4: SUBSCRIPTION PRICE

	Dependent variable: Yearly subscription price							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	-0.273	-0.872	-0.984	-1.103	-1.372	-2.021*	-2.021	-2.781*
	(0.896)	(0.939)	(1.033)	(1.259)	(1.075)	(1.114)	(1.352)	(1.490)
Post-CL × Classified Mgr.					1.966	2.142	1.911	3.403**
					(1.384)	(1.425)	(1.519)	(1.628)
Log population, #ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	19643	19643	19642	18934	19424	19424	19423	18720
Number of newspapers	1431	1431	1431	1387	1393	1393	1393	1349
R^2	0.93	0.94	0.94	0.95	0.93	0.93	0.94	0.94
Mean dependent variable	118.66	118.66	118.67	118.36	118.80	118.80	118.80	118.48

Regressions of yearly subscription price by newspaper and year on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

FIGURE C3: SUBSCRIPTION PRICE – EVENT STUDY



Dynamic effect of the availability of a local CL website on subsription price by newspaper and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

TABLE C5: TOPIC MODEL: WEIGHT ON OTHER TOPICS

	(1) accidents	(2) events	(3) traffic	(4) obituaries	(5) sports	(6) crime	(7) accidents	(8) events	(9) traffic	(10) obituaries	(11) sports	(12) crime
Post-CL	-0.000	-0.009*	-0.005*	0.008	0.014**	-0.000	-0.003	-0.007	-0.004	0.002	0.007	-0.002
	(0.003)	(0.004)	(0.003)	(0.007)	(0.006)	(0.003)	(0.004)	(0.006)	(0.004)	(0.010)	(0.008)	(0.003)
Post-CL × Classified Mgr.							0.006	-0.003	-0.001	0.013	0.012	0.003
							(0.006)	(0.007)	(0.006)	(0.012)	(0.008)	(0.004)
County controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7178	7178	7178	7178	7178	7178	7129	7129	7129	7129	7129	7129
Number of newspapers	869	869	869	869	869	869	862	862	862	862	862	862
R^2	0.44	0.41	0.39	0.56	0.44	0.44	0.43	0.41	0.39	0.56	0.44	0.44
Mean dependent variable	0.12	0.17	0.14	0.15	0.21	0.11	0.12	0.17	0.14	0.15	0.21	0.11

Regressions of topic probabilities aggregated by newspaper and year on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C6: MENTIONS OF NATIONAL PARTY LEADERS

	Dependent variable: Articles mentioning national politicians (IHS)					
	(1)	(2)	(3)	(4)	(5)	(6)
Post-CL	-0.393**	-0.258	-0.195	-0.393**	-0.244	-0.334*
	(0.158)	(0.167)	(0.167)	(0.153)	(0.165)	(0.200)
Post-CL × Classified Mgr.				0.038	-0.017	0.284
				(0.247)	(0.240)	(0.245)
Log population, share urban, # ISPs	Yes	Yes	Yes	Yes	Yes	Yes
Additional county characteristics × Year FEs	No	Yes	Yes	No	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	Yes
Observations	2792	2792	2698	2758	2758	2664
Number of newspapers	518	518	513	512	512	507
R^2	0.71	0.73	0.80	0.71	0.73	0.80
Mean dependent variable	4.51	4.51	4.53	4.50	4.50	4.51

Regressions of the share of articles mentioning the name of a set of national party leaders by newspaper and year on an indicator for the availability of a local Craigslist website, and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C7: SELF-REPORTED NEWSPAPER READERSHIP: HETEROGENEITY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	News propensity				Classified propensity			
	Low		High		Low		High	
Post-CL	-0.011*	-0.010	-0.017***	0.007	-0.023***	-0.005	-0.010	-0.000
	(0.007)	(0.010)	(0.007)	(0.010)	(0.007)	(0.010)	(0.007)	(0.010)
Post-CL × Classified Mgr.		-0.002		-0.032***		-0.024**		-0.013
		(0.012)		(0.011)		(0.012)		(0.012)
Respondent controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	126747	125255	126752	126173	126747	125637	126756	125795
Number of counties								
R^2	0.05	0.05	0.09	0.09	0.11	0.11	0.07	0.07
Mean dependent variable	0.32	0.32	0.52	0.52	0.47	0.47	0.38	0.38

Regressions of self-reported newspaper readership on an indicator for the availability of a local Craigslist website in the county of the respondent, and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. The sample is split by the median respondent-level propensity to read the news section / classified section. Respondent controls include age bins, an indicator for college degree and race. County controls include contemporaneous log population and number of Internet service providers, as well as share urban population, pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share - all measured in 2000 and interacted with year FEs. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C8: SELF-REPORTED MEDIA CONSUMPTION: GFK

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Newspaper, national		TV news		Radio news		Online news	
Post-CL	0.009**	0.007	-0.006	0.004	0.005	0.002	0.003	0.005
	(0.004)	(0.005)	(0.007)	(0.009)	(0.006)	(0.007)	(0.004)	(0.007)
Post-CL × Classified Mgr.		0.003		-0.014		0.004		-0.004
		(0.005)		(0.009)		(0.008)		(0.007)
Respondent controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	253513	251442	253513	251442	253513	251442	253513	251442
Number of counties								
R^2	0.14	0.14	0.08	0.08	0.09	0.09	0.16	0.16
Mean dependent variable	0.09	0.09	0.70	0.70	0.16	0.16	0.22	0.22

Regressions of self-reported media consumption on an indicator for the availability of a local Craigslist website in the county of the respondent, and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. Respondent controls include sex, age, an indicator for college degree and race. County controls include contemporaneous log population and number of Internet service providers, as well as share urban population, pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share - all measured in 2000 and interacted with year FEs. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C9: SELF-REPORTED MEDIA CONSUMPTION: NAES

	(1)	(2)	(3)	(4)	(5)	(6)
	Newspaper, national		TV news		Radio news	
Post-CL	0.030**	-0.000	0.002	0.008	0.004	0.000
	(0.015)	(0.013)	(0.004)	(0.009)	(0.006)	(0.008)
Post-CL × Classified Mgr.		0.037		-0.009		0.005
		(0.024)		(0.009)		(0.008)
Respondent characteristics	Yes	Yes	Yes	Yes	Yes	Yes
County controls	Yes	Yes	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	105989	105173	107778	106981	149111	148083
Number of counties	1207	1203	1207	1203	1208	1204
R^2	0.21	0.20	0.04	0.04	0.05	0.05
Mean dependent variable	0.03	0.03	0.92	0.92	0.37	0.37

Regressions of self-reported media consumption on an indicator for the availability of a local Craigslist website in the county of the respondent, and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. Respondent controls include sex, age, an indicator for college degree and race. County controls include contemporaneous log population and number of Internet service providers, as well as share urban population, pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share - all measured in 2000 and interacted with year FEs. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE C10: SELF-REPORTED MEDIA CONSUMPTION: NAES, CONTINUED

	(1)	(2)	(3)	(4)	(5)	(6)
	Newspaper, national days per wk		TV news days per wk		Radio news days per wk	
Post-CL	0.153**	0.004	0.043	0.094	0.024	0.003
	(0.073)	(0.069)	(0.038)	(0.074)	(0.026)	(0.039)
Post-CL × Classified Mgr.		0.186		-0.068		0.030
		(0.117)		(0.089)		(0.038)
Respondent characteristics	Yes	Yes	Yes	Yes	Yes	Yes
County controls	Yes	Yes	Yes	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	107856	107054	107939	107138	149111	148083
Number of counties	1207	1203	1207	1203	1208	1204
R^2	0.09	0.09	0.10	0.10	0.05	0.05
Mean dependent variable	0.15	0.15	5.04	5.04	1.51	1.51

Regressions of self-reported media consumption on an indicator for the availability of a local Craigslist website in the county of the respondent, and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. Respondent controls include sex, age, an indicator for college degree and race. County controls include contemporaneous log population and number of Internet service providers, as well as share urban population, pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share - all measured in 2000 and interacted with year FEs. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D. Effects for years since CL entry

Table D1: craigslist.org visits; Years since CL entry

	Dependent variable: CL visits (IHS)			
	(1)	(2)	(3)	(4)
Years Post-CL	0.091*	0.004	0.063**	0.051**
	(0.047)	(0.038)	(0.027)	(0.023)
Total Comscore visits (IHS)	Yes	Yes	Yes	Yes
Log population, # ISPs	Yes	Yes	Yes	Yes
2000 county char. × Year FEs	No	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes
Observations	20396	20396	20396	20007
Number of counties	3043	3043	3043	2986
R^2	0.79	0.82	0.83	0.85
Mean dependent variable	2.36	2.36	2.36	2.34

Regressions of number of visits of the domain craigslist.org by county and year (IHS-transformed) on an indicator for the availability of a local Craigslist website. All specifications control for total visits recorded in Comscore (IHS-transformed). County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE D2: NUMBER OF NEWSPAPERS; YEARS SINCE CL ENTRY

	Dependent variable: Number of newspapers HQ-ed in county							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years Post-CL	0.002	-0.001	-0.002	-0.002	0.007	0.003	0.001	-0.001
	(0.006)	(0.005)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.008)
Years Post-CL × Classified Mgr.					-0.009	-0.007	-0.006	-0.003
					(0.008)	(0.008)	(0.008)	(0.008)
Log population, # ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × Year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	19424	19424	19424	18672	19360	19360	19360	18608
Number of counties	1214	1214	1214	1167	1210	1210	1210	1163
R^2	0.94	0.95	0.95	0.95	0.94	0.95	0.95	0.95
Mean dependent variable	1.18	1.18	1.18	1.18	1.18	1.18	1.18	1.18

Regressions of the number of newspapers HQ-ed by county and year on an indicator for the availability of a local Craigslist website and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE D3: NUMBER OF JOBS; YEARS SINCE CL ENTRY

	Dependent variable: Newspaper number of jobs							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years Post-CL	-0.833***	-0.513***	-0.553***	-0.539***	-0.125	0.126	0.080	0.061
	(0.163)	(0.152)	(0.169)	(0.194)	(0.102)	(0.109)	(0.124)	(0.162)
Years Post-CL × Classified Mgr.					-1.164***	-1.099***	-1.090***	-1.037**
					(0.216)	(0.206)	(0.203)	(0.213)
Log population, #ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	22724	22724	22723	21968	22516	22516	22515	21762
Number of newspapers	1540	1540	1540	1492	1505	1505	1505	1459
R^2	0.90	0.91	0.91	0.92	0.91	0.91	0.91	0.92
Mean dependent variable	21.31	21.31	21.31	21.14	21.38	21.38	21.39	21.23

Regressions of number of jobs by newspaper and year on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

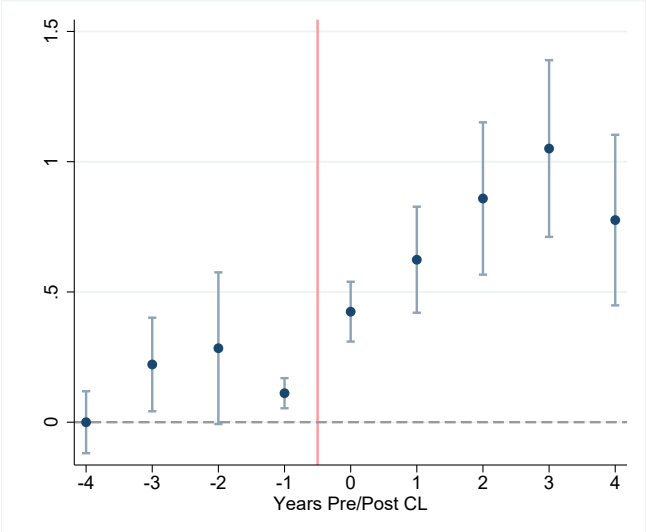
TABLE D4: CIRCULATION; YEARS SINCE CL ENTRY

	Dependent variable: Newspaper circulation in thousands							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Years Post-CL	-2.416***	-1.483***	-1.494***	-1.760***	-1.298***	-0.669	-0.684	-0.990*
	(0.637)	(0.567)	(0.529)	(0.438)	(0.349)	(0.441)	(0.436)	(0.517)
Years Post-CL × Classified Mgr.					-1.912**	-1.467*	-1.433*	-1.398
					(0.800)	(0.774)	(0.779)	(0.911)
Log population, #ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	22950	22950	22948	22195	22633	22633	22631	21880
Number of newspapers	1555	1555	1555	1507	1506	1506	1506	1460
R^2	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.99
Mean dependent variable	33.34	33.34	33.34	32.89	33.40	33.40	33.39	32.93

Regressions of circulation by newspaper and year, measured in thousands of copies, on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

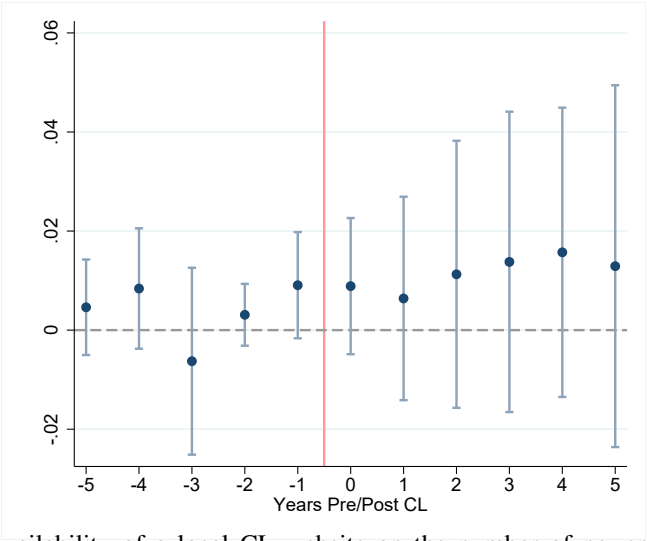
E. "Broad" definition of CL and newspaper markets

Figure E1: craigslist.org visits – Event Study; Broad CL and newspaper markets



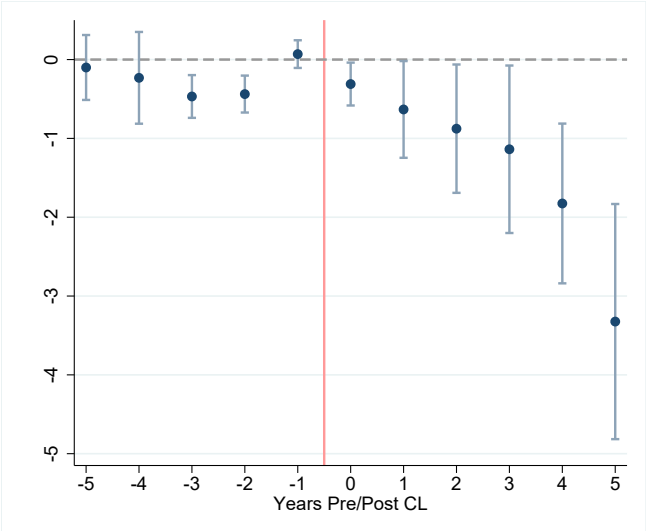
Dynamic effect of the availability of a local CL website on the number of craigslist.org visits by county and year (IHS-transformed). Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include total Comscore visits (IHS transformed), log population and number of Internet service providers. Standard errors clustered by CL-area.

Figure E2: Number of newspapers – Event Study; Broad CL and newspaper markets



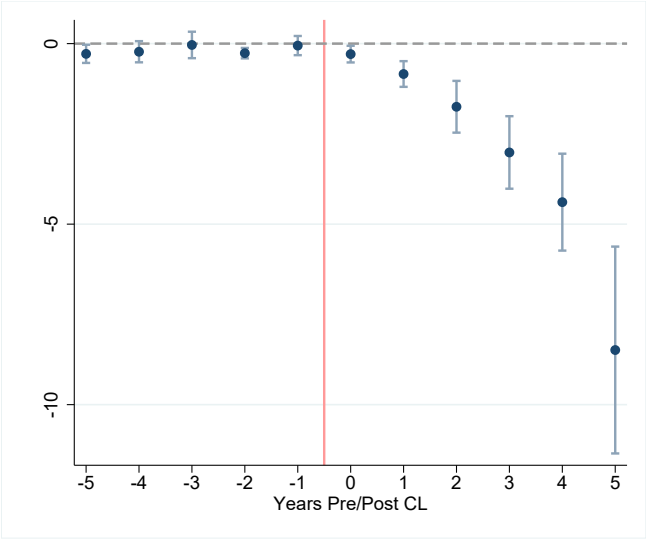
Dynamic effect of the availability of a local CL website on the number of newspapers HQ-ed by county and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

FIGURE E3: NUMBER OF JOBS – EVENT STUDY; BROAD CL AND NEWSPAPER MARKETS



Dynamic effect of the availability of a local CL website on number of jobs by newspaper and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

FIGURE E4: CIRCULATION – EVENT STUDY; BROAD CL AND NEWSPAPER MARKETS



Dynamic effect of the availability of a local CL website on circulation (in thousands of copies) by newspaper and year. Coefficients and 95% confidence intervals based on the time-corrected Wald estimator proposed in de Chaisemartin and D'Haultfoeuille (2020). Controls include log population and number of Internet service providers. Standard errors clustered by CL-area.

TABLE E1: CRAIGSLIST.ORG VISITS; BROAD CL AND NEWSPAPER MARKETS

	Dependent variable: CL visits (IHS)			
	(1)	(2)	(3)	(4)
Post-CL	0.533***	0.226***	0.208***	0.191***
	(0.103)	(0.067)	(0.060)	(0.051)
Total Comscore visits	Yes	Yes	Yes	Yes
Log pop., pct. urban, # ISPs	Yes	Yes	Yes	Yes
2000 county char. × Year FEs	No	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes
Observations	20396	20396	20396	20007
Number of counties	3043	3043	3043	2986
R^2	0.79	0.82	0.83	0.85
Mean dependent variable	2.36	2.36	2.36	2.34

Regressions of number of visits of the domain craigslist.org by county and year (IHS-transformed) on an indicator for the availability of a local Craigslist website. All specifications control for total visits recorded in Comscore (IHS-transformed). County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE E2: NUMBER OF NEWSPAPERS – BROAD CL AND NEWSPAPER MARKETS

	Dependent variable: Number of newspapers HQ-ed in county							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	0.029***	0.023**	0.017	0.016	0.036***	0.026**	0.022*	0.015
	(0.011)	(0.011)	(0.011)	(0.014)	(0.012)	(0.012)	(0.012)	(0.015)
Post-CL × Classified Mgr.					-0.012	-0.004	-0.010	0.002
					(0.015)	(0.015)	(0.015)	(0.016)
Log pop., pct. urban, # ISPs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2000 county char. × Year FEs	No	Yes	Yes	Yes	No	Yes	Yes	Yes
County FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	19200	19200	19200	18416	19152	19152	19152	18368
Number of counties	1200	1200	1200	1151	1197	1197	1197	1148
R^2	0.95	0.95	0.95	0.96	0.95	0.95	0.95	0.96
Mean dependent variable	1.19	1.19	1.19	1.19	1.19	1.19	1.19	1.19

Regressions of the number of newspapers HQ-ed by county and year on an indicator for the availability of a local Craigslist website and its interaction with the circulation-weighted share of newspapers with a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE E3: NUMBER OF JOBS – BROAD CL AND NEWSPAPER MARKETS

	Dependent variable: Newspaper number of jobs							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	-1.200***	-0.727**	-0.924***	-1.280***	0.582	0.750**	0.555	0.290
	(0.364)	(0.313)	(0.341)	(0.381)	(0.392)	(0.348)	(0.347)	(0.375)
Post-CL × Classified Mgr.					-3.852***	-3.249***	-3.272***	-3.572***
					(0.544)	(0.529)	(0.526)	(0.567)
County controls #1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls #2	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	22418	22418	22402	21611	22382	22382	22366	21575
Number of newspapers	1454	1454	1453	1403	1450	1450	1449	1399
R^2	0.90	0.91	0.91	0.91	0.90	0.91	0.91	0.91
Mean dependent variable	21.50	21.50	21.48	21.24	21.50	21.50	21.49	21.25

Regressions of number of jobs by newspaper and year on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE E4: TOPIC MODEL: WEIGHT ON POLITICAL TOPICS; BROAD CL AND NEWSPAPER MARKETS

	(1) president	(2) local	(3) congress	(4) foreign	(5) president	(6) local	(7) congress	(8) foreign
Post-CL	-0.007*	-0.011**	-0.005*	0.003	0.000	-0.007	-0.001	0.010***
	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)	(0.006)	(0.003)	(0.003)
Post-CL × Classified Mgr.					-0.015**	-0.008	-0.008*	-0.013***
					(0.006)	(0.007)	(0.004)	(0.004)
Full county controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7178	7178	7178	7178	7129	7129	7129	7129
Number of newspapers	869	869	869	869	862	862	862	862
R^2	0.52	0.47	0.41	0.54	0.52	0.47	0.40	0.54
Mean dependent variable	0.21	0.31	0.10	0.10	0.21	0.31	0.10	0.10

Regressions of the average probability of an article covering a particular topic by newspaper and year on an indicator for the availability of a local Craigslist website, and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

TABLE E5: CIRCULATION – BROAD CL AND NEWSPAPER MARKETS

	Dependent variable: Newspaper circulation in thousands							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post-CL	-1.161*	-1.058	-1.288*	-0.827	1.348*	0.526	0.236	0.750
	(0.620)	(0.726)	(0.709)	(0.741)	(0.762)	(0.678)	(0.619)	(0.749)
Post-CL × Classified Mgr.					-5.407***	-3.428***	-3.261***	-3.421***
					(1.429)	(1.122)	(1.083)	(1.036)
County controls #1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County controls #2	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Newspaper FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × Year FEs	No	No	Yes	No	No	No	Yes	No
DMA × Year FEs	No	No	No	Yes	No	No	No	Yes
Observations	22590	22590	22574	21784	22497	22497	22481	21691
Number of newspapers	1457	1457	1456	1406	1450	1450	1449	1399
R^2	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
Mean dependent variable	34.12	34.12	34.10	33.11	34.11	34.11	34.09	33.09

Regressions of circulation by newspaper and year, measured in thousands of copies, on an indicator for the availability of a local Craigslist website and its interaction with an indicator for the presence of a classified manager at baseline. County characteristics for the year 2000 include pct. college educated, pct. rental, median age, share white/ black/ hispanic, income per capita, unemployment rate, presidential turnout and Republican vote share. OLS regressions in all columns. Standard errors clustered by CL-area. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.