NBER WORKING PAPER SERIES

# WHAT DRIVES MEDIA SLANT? EVIDENCE FROM U.S. DAILY NEWSPAPERS

Matthew Gentzkow
Jesse M. Shapiro

Working Paper 12707
http://www.nber.org/papers/w12707

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2006

We are grateful to Gary Becker, Gary Chamberlain, Raj Chetty, Tim Conley, Edward Glaeser, Tim Groseclose, Christian Hansen, Larry Katz, Kevin M. Murphy, Andrea Prat, Sam Schulhofer-Wohl, Andrei Shleifer, Wing Suen, Abe Wickelgren, and seminar participants at the University of Chicago, the University of Illinois at Chicago, Clemson University, the University of Toronto, the University of Houston, Rice University, Texas A&M University, Harvard University, and Northwestern University for helpful comments. We especially wish to thank Renata Voccia, Paul Wilt, Todd Fegan, and the rest of the staff at ProQuest for their support and assistance at all stages of this project. Steve Cicala, Hays Golden, Jennifer Paniza, and Mike Sinkinson provided outstanding research assistance and showed tireless dedication to this project. We also thank Yujing Chen, Alex Fogel, and Lisa Furchtgott for excellent research assistance. This research was supported by National Science Foundation Grant SES-0617658. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

What Drives Media Slant? Evidence from U.S. Daily Newspapers
Matthew Gentzkow and Jesse M. Shapiro
NBER Working Paper No. 12707
November 2006
JEL No. D78,K23,L82


## ABSTRACT

We construct a new index of media slant that measures whether a news outlet's language is more similar to a congressional Republican or Democrat. We apply the measure to study the market forces that determine political content in the news. We estimate a model of newspaper demand that incorporates slant explicitly, estimate the slant that would be chosen if newspapers independently maximized their own profits, and compare these ideal points with firms' actual choices. Our analysis confirms an economically significant demand for news slanted toward one's own political ideology. Firms respond strongly to consumer preferences, which account for roughly 20 percent of the variation in measured slant in our sample. By contrast, the identity of a newspaper's owner explains far less of the variation in slant, and we find little evidence that media conglomerates homogenize news to minimize fixed costs in the production of content.



Matthew Gentzkow
University of Chicago
Graduate School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
gentzkow@chicagogsb.edu

Jesse M. Shapiro
University of Chicago
5807 S. Woodlawn Avenue
Chicago, IL 60637
and NBER
jmshapir@uchicago.edu

# 1 Introduction

Government regulation of news media ownership in the United States is built on two propositions. The first is that news content has a powerful impact on politics, with ideologically diverse content producing socially desirable outcomes. According to the U.S. Supreme Court: "One of the most vital of all general interests [is] the dissemination of news from as many different sources, and with as many different facets and colors as is possible. That interest...presupposes that right conclusions are more likely to be gathered out of a multitude of tongues, than through any kind of authoritative selection" (U.S. Supreme Court 1945).[1]

The second proposition is that unregulated markets will tend to produce too little ideological diversity. The Federal Communications Commission (FCC), for example, "has traditionally assumed that there is a positive correlation between viewpoints expressed and ownership of an outlet. The Commission has sought, therefore, to diffuse ownership of media outlets among multiple firms in order to diversify the viewpoints available to the public" (FCC 2003).[2] This belief has justified significant controls on cross-market consolidation in broadcast media ownership, on foreign ownership of media, and on cross-media ownership within markets, and has motivated a sizable academic literature arguing that current media ownership is too concentrated (Bagdikian 2000).

That news content can have significant effects on political attitudes and outcomes has been documented empirically by Strömberg (2004), Gentzkow and Shapiro (2004), Gentzkow (2006), DellaVigna and Kaplan (2007), Gerber, Karlan, and Bergan (2006), and others. In contrast, evidence on the way market forces interact to produce ideological content, and on the role of ownership in particular, is severely limited. Existing studies have generally relied on hand collection and coding of news content, and so have been limited to analysis of a few outlets (e.g., Glasser, Allen, and Blanks 1989; Pritchard 2002). Groseclose and Milyo (2005) make an important contribution, proposing a new measure of ideological content based on counts of think-tank citations. However,

[1] The Federal Communications Commission (2003) echoes the same point: "Viewpoint diversity refers to the availability of media content reflecting a variety of perspectives. A diverse and robust marketplace of ideas is the foundation of our democracy. Consequently, 'it has long been a basic tenet of national communications policy that the widest possible dissemination of information from diverse and antagonistic sources is essential to the welfare of the public.'"

[2] The report of the Hutchins Commission (Commission on Freedom of the Press 1947), arguably the most influential study of public policy and the press, identifies the "first and foremost" obstacle to the emergence of truth in the press as "the drift toward concentration of power...exemplified by the large number of cities with only one newspaper, the common ownership of newspapers and radio stations, and the growth of newspaper chains."

their index has been calculated only for a small number of outlets, and has not been used to analyze the determinants of slant.

In this paper, we propose a new index of ideological slant in news coverage, and compute it for a large sample of U.S. daily newspapers. We then apply the measure to study the way market forces determine slant in equilibrium. We estimate a model of newspaper demand that incorporates slant explicitly, estimate the slant that would be chosen if newspapers independently maximized their own profits, and compare these ideal points with firms' actual choices. Finally, we use the model to evaluate the contributions of consumer and owner heterogeneity to ideological diversity, and test a variety of other theories of what drives media slant.

Our slant index measures the frequency with which newspapers use language that would tend to sway readers to the right or to the left on political issues.[3] To do this, we examine the set of all phrases used by members of Congress in the 2005 *Congressional Record*, and identify those that are used much more frequently by one party than by another. We then index newspapers by the extent to which the use of politically charged phrases in their news coverage resembles the use of the same phrases in the speech of a congressional Democrat or Republican. Underlying this approach is a revealed preference assumption; namely, that the language chosen by speakers with a political agenda will tend to persuade listeners to support that agenda.

Two key pieces of evidence suggest that our methodology produces a meaningful measure of slant. First, many of the phrases that our automated procedure identifies are known from other sources to be chosen strategically by politicians for their persuasive impact. Examples include "death tax," "tax relief," "personal account," and "war on terror" (which we identify as strongly Republican), and "estate tax," "tax break," "private account," and "war in Iraq," (which we identify as strongly Democratic). Second, the index that we construct using counts of these phrases in news coverage is consistent with readers' subjective evaluation of newspapers' political leanings (data on which is available for several large papers in our sample), and with estimates in Groseclose and Milyo (2005).

We use our measure to estimate a Hotelling model of newspaper demand, in which a con-

[3] The term "slant" was apparently introduced by Hayakawa (1942). He uses it to refer to the process of creating an impression through selective omission or inclusion of facts. We use the term more inclusively to include any differences in news content that, *ceteris paribus*, would tend to increase a reader's support for one side of the political spectrum or the other.

sumer's utility from reading a newspaper depends on the match between the newspaper's slant and the consumer's own ideology (Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006). Using zipcode-level data on newspaper circulation, we show that right-wing newspapers circulate relatively more in zipcodes with high fractions of Republicans, even within a narrowly defined geographic market. This fact is robust to correcting for the endogeneity of newspaper slant using an identification strategy in the spirit of George and Waldfogel (2003), and survives a specification with zipcode fixed effects. Our model implies that consumer demand for slant is not only statistically but also economically significant; a one-standard-deviation move away from the profit-maximizing level of slant would lead to a loss in circulation of approximately 3.4 percent. Our model also allows us to predict the slant that each newspaper would choose if it independently maximized its own profits, as a function of the share of Republicans in its market.

We next turn to the supply-side of the market, comparing our estimates of profit-maximizing slant to the actual slant chosen by newspapers. Our first finding is that newspapers' actual slant is close to the profit-maximizing level on average. Consistent with Groseclose and Milyo (2005), we find that the average newspaper's language is similar to that of a left-of-center member of Congress. However, we estimate that the profit-maximizing points are also left-of-center on average, perhaps because of demographic differences between readers and non-readers of newspapers, and we cannot statistically reject that the distributions of actual and ideal slant have the same mean. These findings have direct relevance for theories that posit an aggregate bias in the news, such as a bias in the direction of reporters' preferences (Baron 2006), or in favor of incumbent politicians (Besley and Prat 2006).

Our second finding is that the variation in slant across newspapers is strongly related to our estimated ideal points. The relationship remains when we compare different newspapers with the same owner, or different newspapers located in the same state. Moreover, predetermined characteristics, such as the religiosity of the local population, are highly predictive of newspaper slant, indicating that the effects we estimate are not driven by reverse causality from newspaper content to political beliefs. We further argue that variation in the tastes of reporters and editors across local markets is unlikely to be driving the relationship we observe. Overall, the variation in consumer political attitudes incorporated in our estimated ideal points explains roughly 20 percent

of the variation in measured slant in our sample.

Our third finding is that the role of ownership in determining slant is significantly smaller than the role of consumer characteristics. After controlling for geographic clustering of newspaper ownership groups, the slant of co-owned papers is only a weak (and statistically insignificant) predictor of a newspaper's political alignment. Estimates from a random effects model suggest that ownership does not account for any of the variation in measured slant, with a confidence interval that rules out economically large effects. Direct proxies for owner ideology, such as patterns of corporate or executive donations to political parties, are unrelated to a newspaper's slant. We also find that newspapers do not cater to the preferences of consumers in markets served by co-owned newspapers, casting doubt on the hypothesis that fixed costs in the production of news lead to homogenization of slant within conglomerates.

This paper presents some of the first direct empirical evidence on the determinants of political slant in the news.[4] Our findings on the demand for slant, and on newspapers' response to consumer preferences, support theories that posit a role for consumers' prior beliefs in driving media positioning (e.g., Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006; Suen 2004). Our findings on the average slant in the media, the role of owner ideology, and the response of slant to political incentives inform models that focus on the role of the suppliers of news in determining its content (e.g., Besley and Prat 2006; Balan, DeGraba, and Wickelgren 2005; Baron 2006).

Our work also advances the measurement of media slant (Groseclose and Milyo 2005; Puglisi 2006; Lott and Hassett 2004; Gentzkow, Glaeser, and Goldin 2006).[5] Groseclose and Milyo (2005) use Congressional citations to estimate the political positions of think tanks, and then use data on media mentions of the same set of think tanks to measure the bias of 20 news outlets. Our automated procedure allows us to measure the slant of a much wider range of outlets, including

[4] An existing literature explores the determinants of newspaper endorsements of political candidates, rather than news content (see, e.g., Akhavan-Majid, Rife, and Gopinath 1991; or Ansolabehere, Lessem, and Snyder 2006). We focus on news content because of its centrality to public policy debates, and because it is likely to exhibit very different variation across newspapers. Indeed, in specifications not reported in the paper, we find that a considerable portion of the variation in slant is independent of endorsements, and that, after controlling for news slant, consumer demand does not depend on endorsements. These findings imply that news and editorial slant are very different both statistically and economically, suggesting that our emphasis on news content is likely to reveal important patterns not visible when using data on endorsements.

[5] Our approach borrows tools from the computer science literature on "text categorization" (see Aas and Eikvil 1999 for a review), which social scientists have applied to the measurement of sentiment (e.g., Antweiler and Frank 2004), and politicians' platforms (Laver, Benoit, and Garry 2003), but not to the political slant of the news media.

over 400 daily newspapers representing over 70 percent of total daily circulation in the United States. Moreover, rather than imposing a list of likely partisan phrases (such as names of think tanks), we use data from Congress to isolate the phrases that have the most power to identify the speaker's ideology. This methodology is likely to increase precision. It is also applicable in situations in which a list of politically slanted sources is not available, or when sources such as think tanks are rarely cited.

Finally, our findings contribute to the literature on product positioning in the mass media (Sweeting 2006; Myers 2005; George 2001), as well as to research on product differentiation more generally (Mazzeo 2002a and 2002b; Dranove, Gron, and Mazzeo 2003; Seim forthcoming). The existence of rich, within-market variation in consumer ideology allows us to estimate the demand for slant without assuming that it is chosen optimally. We can then use variation across markets to test the hypothesis that slant is chosen to maximize profits. With a few exceptions (e.g., Dubé, Hitsch, and Manchanda 2005; Einav forthcoming), data limitations have generally made such comparisons difficult.

The remainder of the paper is organized as follows. Section 2 discusses our data sources, including information on the databases we use to count phrase frequencies in the *Congressional Record* and in newspapers. Section 3 describes the computation of our measure of newspaper slant, and shows evidence validating this measure against alternative rankings of newspapers' political content. Section 4 discusses our model and estimates of the demand for slant. Section 5 presents theory and evidence on the supply of slant. Section 6 presents additional evidence on the mechanisms driving slant, testing for fixed costs in the production of news and exploiting variation in financial and political incentives. Section 7 concludes.

# 2 Data

## 2.1 *Congressional Record* and Congressperson Characteristics

Our approach to measuring slant requires data on the frequency with which individual members of Congress use particular phrases. We use the complete text of the 2005 *Congressional Record*,

parsed using an automated script that identifies the speaker of each passage.[6]

To increase the efficiency of our text analysis algorithm, we apply a standard pre-processing procedure that removes extremely common words (such as "to," "from," and "the") and strips words down to shared linguistic roots (so that, for example, "tax cut" and "tax cuts" are identified as the same phrase).[7] A final script produces counts by speaker and party of two- and three-word phrases in the *Congressional Record*.[8]

For each congressperson,[9] we obtain data on party identification, as well as the share of the 2004 two-party presidential vote total going to George W. Bush in the congressperson's constituency (congressional district for representatives; state for senators). This vote share serves as our primary measure of a congressperson's ideology, and is strongly correlated with voting behavior as measured by the congressperson's adjusted ADA score (Groseclose, Levitt, and Snyder 1999).[10] We also obtain data on the state, Census division, and demographic characteristics of each congressperson's constituency.[11]

## 2.2 Newspaper Text and Characteristics

As an input to our slant measure, we obtain counts of the frequency with which phrases appear in news coverage from two sources: the NewsLibrary database (<http://www.newslibrary.com>) and the ProQuest Newsstand database (<http://proquest.com>). For each database, we use an

[6] We use an automated script to download the *Congressional Record* from <http://thomas.loc.gov/>. We wish to focus on floor speeches rather than text that is primarily procedural, so we exclude speech by officers such as the Clerk, the Speaker of the House, and the President of the Senate. We also exclude block quotations, text that is inserted into the Record from other sources such as reports or letters, and non-speech items like records of roll-call votes.

[7] We used a list of extremely common words ("stopwords") from Fox (1990). We use the "Porter Stemmer" (<http://www.tartarus.org/martin/PorterStemmer/>) to strip words down to their linguistic roots. We also remove several proper nouns that appear frequently in procedural text—days of the week, the "Hart Senate Office Building," and the "Dirksen Senate Office Building." Finally, we exclude names of major newspapers.

[8] We exclude single words because they occur with higher frequency than phrases and so are costly to search for in newspapers. Preliminary investigation also suggested that most single words are used across many contexts and so generate a high noise to signal ratio. Phrases of four or more words are costly in terms of text processing, and in a preliminary analysis did not appear to add significantly to the precision of our measure.

[9] We use the word "congressperson" as a generic term to refer to members of both the House of Representatives and the Senate.

[10] The correlation coefficient is $-.75$ (higher ADA scores correspond to more liberal politicians). We thank Tim Groseclose for providing us with adjusted ADA scores covering members of congress through 1999. Because our analysis is based on the 2005 *Congressional Record*, the correlation coefficient is for the sub-sample of members who were present in 1999.

[11] Data on presidential vote shares and demographic characteristics of congressional districts are from <www.polidata.org>.

automated script to calculate the number of articles containing each phrase in each newspaper during calendar year 2005. Whenever possible, we exclude opinion and editorial content. Also, because some newspapers do not archive reprinted wire stories with ProQuest, we exclude articles from the Associated Press and other national wire stories, focusing instead on content originating with the newspaper.

We include in our sample only English-language, daily newspapers. Data are available for 377 such newspapers from NewsLibrary (NL) and 155 from ProQuest (PQ), with an overlap of 115 newspapers. This leaves us with a total sample of 417 newspapers in 2005.[12] Among the newspapers that overlap between the two databases, the correlation between the counts for the sample of phrases described below is over .80, indicating high cross-database reliability.[13] In cases of overlap, we use the NewsLibrary counts for analysis.

To measure the ownership and market characteristics of the newspapers in our sample, we first match every newspaper to data from the 2000 Editor and Publisher (E&P) International Yearbook CD-ROM. The E&P dataset identifies the zipcode of each newspaper's headquarters, which we match to counties using the United States 5-Digit ZIP Code Database from Quentin Sager Consulting. We match counties to primary metropolitan statistical areas (PMSAs) using definitions from the 1990 census. We define each newspaper's geographic market as the PMSA in which it is headquartered. If a newspaper is not located inside a PMSA, we define its market to be the county in which it is located.

For each newspaper, we obtain, from the 2000 U.S. Census, a wide range of demographic characteristics of the paper's market. We also obtain data from David Leip's Atlas of US Presidential Elections (<http://www.uselectionatlas.org>) on the share of votes in each market going to Bush in the 2004 presidential election, as a proxy for the market's political leanings. Lastly, we use

[12] One additional newspaper—the *Chicago Defender*—is present in the news databases, but is excluded from our analysis because it is an extreme outlier (more than 13 standard deviations away from the mean) in the distribution of slant. The vast majority of hits for this paper are for a single phrase, "African American," which is strongly predictive of liberal ideology in Congress.

[13] There are several reasons why there is not perfect agreement between the two databases. The first is that there is a lag between the publication of an article and its posting to the news database. Although we began our searches several months after the end of 2005, and attempted to search both PQ and NL simultaneously, it is still possible that some news stories occurring late in the year were not catalogued at the time of the PQ search and were catalogued at the time of the NL search, or vice versa. The second is that there are some small differences in how the two databases permit us to identify editorials and opinion pieces, as well as wire stories, although hand-checking suggests this issue is less significant than the posting lag in explaining the discrepancies between the two databases.

the DDB Needham Life Style Survey (Putnam 2000), available on <www.bowlingalone.com>, to compute a measure of the share of survey respondents from 1972-1998 who report attending church monthly or more. This measure serves as a plausibly exogenous shifter of the political leanings of the market in that it is unlikely to be directly affected by the slant of area newspapers.

The E&P dataset provides information on a number of newspaper characteristics, such as the number of pages in the paper and the number of employees, which serve as a proxy for the quality of the newspaper (Berry and Waldfogel 2003). We also obtain data from <www.pulitzer.org> on the number of Pulitzer prizes won by each newspaper since 1970. The E&P dataset identifies the owner of each newspaper as of 2000, which we augment with information on ownership from 2001-2005 from the hard-copy *Editor and Publisher International Yearbook.*

As a potential proxy for a media firm's ideological leanings, we obtain data from the Center for Public Integrity (<http://www.publicintegrity.org>) on the share of each newspaper owner's corporate political contribution dollars going to Republicans. We also searched the Federal Election Commission (FEC) disclosure database for information on the personal contributions of the CEO, President, Chairman, and Managing Director of each firm that owns two or more newspapers, and computed an aggregate measure of the share of donation dollars going to Republicans.

### 2.3 Newspaper Circulation and Consumer Characteristics

For our study of the effects of slant on newspaper demand, we use zipcode-level data on newspaper circulation from the Audit Bureau of Circulation's (ABC) Newspaper GeoCirc dataset, which covers 290 of the papers in our sample. We match each zipcode to a news market. If a given market contains no readers of a newspaper, we exclude observations in that market-newspaper pair from the dataset.[14]

To adjust for non-political differences across zipcodes, we make use of a set of zipcode demographics taken from the 2000 U.S. Census (<www.census.gov>).

Measuring each zipcode's political preferences is complicated by the fact that voting data are not available at the zipcode level. To circumvent this problem, we use the Federal Election Commission's (FEC) 2000, 2002, and 2004 Individual Contributions Files. These files, which are available for

[14] Because our analysis will use only variation across zipcodes within a market, such cases provide no additional variation with which to identify our models.

download at <http://www.fec.gov>, contain a record of every individual contribution to a political party, candidate, or political action committee registered with the FEC. Each donor record includes a complete address, allowing us to identify donors' zipcodes. For each zipcode, we compute the share of donations received by a Republican affiliate, among donations received by either Republican- or Democrat-affiliated entities.

This calculation gives us a noisy, but informative proxy for the political attitudes of each zipcode. Although we cannot compare our donation measure with voting information by zipcode, we can compare the share of donations to Republicans at the county level with the county's Republican vote share in 2000. Our donation measure is available for over 99.5% of counties in the U.S., and has a highly statistically significant correlation of 0.40 with Bush's share of the two-party vote in 2000 ($p < 0.0001$).

Of course, the sample of donors to political causes is not fully representative of the entire population of a zipcode. Donors tend to be older, richer, and more educated than non-donors (Gimpel, Lee, and Kaminski forthcoming). However, these are also the demographic characteristics of likely readers of newspapers (Gentzkow forthcoming), and therefore, if anything, may tend to make our measure more representative of the population relevant for studying newspaper demand.

## 3 Measuring Slant

Our approach to measuring the slant of a newspaper will be to compare phrase frequencies from the newspaper with phrase frequencies in the 2005 *Congressional Record*, in order to identify whether the newspaper's language is more similar to that of a congressional Republican or a congressional Democrat. Following a large literature in computer science on "text categorization" (Aas and Eikvil 1999), we proceed in two steps. First, we select a subset of the millions of phrases in the *Congressional Record* to use for our analysis. Second, we aggregate the frequencies of the resulting phrases into a single measure of political slant.

For a concrete illustration of our approach to measuring slant, consider the use of the phrases "death tax" and "estate tax" to describe the federal tax on assets of the deceased. The phrase "death tax" was coined by the tax's conservative opponents. According to a high-level Republican

staffer, "Republicans put a high level of importance on the death/estate tax language—they had to work hard to get members to act in unison, including training members to say 'death tax'... Estate tax sounds like it only hits the wealthy but 'death tax' sounds like it hits everyone" (Graetz and Shapiro 2005). In the U.S. House of Representatives in 2005, Republicans used the phrase "death tax" 365 times and the phrase "estate tax" only 46 times. Democrats, by contrast, had the reverse pattern, using the phrase "death tax" only 35 times and the phrase "estate tax" 195 times.

The relative use of the two phrases in newspaper text conforms well to prior expectations about political slant. Compare, for example, the *Washington Times* and the *Washington Post.* The former is widely perceived to be a conservative newspaper, while the latter is generally thought to be more liberal.[15] In 2005, the *Post* used the phrase "estate tax" 10 times more often than it used the phrase "death tax," while the *Times* used the phrase "estate tax" only twice as often. As we show below, this case is not unusual: there is a strong correlation between popular perceptions of a newspaper's political leanings and its propensity to use words and phrases favored by different political parties in Congress. Our measure of media slant exploits this fact by endogenously identifying politically charged phrases like "death tax" and "estate tax," and computing their frequencies in daily newspapers throughout the United States.

### 3.1 Selecting Phrases for Analysis

In order to make the analysis manageable, we first need to select from the millions of phrases that appear in congressional speech a subset of phrases that are likely to be informative about partisanship. To do so, we measure the extent to which each phrase is used differentially by one party or the other. Let $f_{pd}$ and $f_{pr}$ denote the total number of times phrase $p$ is used by Democrats and Republicans respectively. Let $f_{\sim pd}$ and $f_{\sim pr}$ denote the total count of phrases that are *not* phrase $p$ spoken by Democrats and Republicans, respectively (where we restrict attention to the set of phrases with the same number of words as $p$). To identify partisan phrases, we compute a Pearson's $\chi^2$ statistic for the null hypothesis that the propensity to use phrase $p$ is equal for

[15] The website <www.mondotimes.com> presents an index of newspapers' political leanings based on user ratings. The *Times* is rated as "conservative" while the *Post* is rated as "leans left." Groseclose and Milyo (2005) also rate the *Times* as significantly to the right of the *Post.*

Democrats and Republicans:

$$\chi_p^2 = \frac{\left(f_{pr} f_{\sim pd} - f_{pd} f_{\sim pr}\right)^2}{\left(f_{pr} + f_{pd}\right)\left(f_{pr} + f_{\sim pr}\right)\left(f_{pd} + f_{\sim pd}\right)\left(f_{\sim pr} + f_{\sim pd}\right)} \tag{1}$$

The $\chi^2$ statistic is a convenient summary of the political asymmetry in the use of a phrase, because it incorporates both how often the phrase is used by each party and its overall importance in political speeches. (More naive statistics, such as the ratio of uses by Republicans to uses by Democrats, would tend to select phrases that are used only once by Republicans and never by Democrats, even though pure sampling error could easily generate such a pattern.) It is also simple to compute, in the sense that it requires only two calculations per phrase: the number of uses by Republicans, and the number of uses by Democrats.

In addition to the $\chi^2$ statistic, we also compute the total number of times that each phrase appeared in newspaper headlines and article text in the ProQuest Newsstand database from 2000-2005. In order to be useful for our purposes, a phrase must be in sufficiently common use to actually show up routinely in newspaper searches. Procedural phrases, such as "yield the remainder of my time," which are commonly employed in the *Congressional Record* but are almost never used outside of parliamentary contexts, are unlikely to be helpful in identifying the slant of a newspaper.[16] Additionally, phrases that are extremely common, such as "third quarter" or "exchange rate" would generate a large number of hits and so have a high computational cost relative to the additional information they convey. We therefore restrict attention to two-word phrases that appeared in at least 200 but no more than $15,000$ newspaper headlines, and three-word phrases that appeared in at least 5 but no more than $1,000$ headlines. We also drop any phrase that appeared in the full text of more than $400,000$ documents.[17] Our final set consists of the top 500 two-word and top 500 three-word phrases by $\chi^2$ that satisfy this criterion, for a total of 1,000 phrases.

Table 1 shows the top phrases (by $\chi^2$) in our final set of 1,000.[18] The first panel shows phrases

[16] Parliamentary protocol means that a number of procedural phrases are used more often by either the majority or minority party, and so show up as partisan speech according to the $\chi^2$ measure.

[17] These cutoffs are arbitrary. They were chosen to exclude as efficiently as possible both procedural text (on the bottom end) and extremely common everyday phrases (on the top end). When we tighten the cutoffs by excluding, for example, the top and bottom five percent of phrases ranked by the total number of headlines mentioning the phrase, the resulting measure is highly correlated with our own, and produces similar results statistically. (See appendix A for details.) Our findings therefore do not seem particularly sensitive to the choice of headline count cutoffs for these phrases.

[18] Casual inspection reveals that some of the two-word phrases on our list are proper subsets of three-word phrases

used more often by congressional Democrats. The second panel shows phrases used more often by congressional Republicans.

Our procedure identifies many phrases that both intuition and existing evidence suggest are chosen strategically for their partisan impact. For example, a widely circulated 2005 memo by Republican consultant Frank Luntz advised candidates on the language they should use to describe President Bush's proposed social security reform:

> Never say 'privatization/private accounts.' Instead say 'personalization/personal accounts.' Two-thirds of America want to personalize Social Security while only one-third would privatize it. Why? Personalizing Social Security suggests ownership and control over your retirement savings, while privatizing it suggests a profit motive and winners and losers (Luntz 2005).

We identify "personal accounts," "personal retirement accounts," and "personal savings accounts" as among the most Republican phrases in the *Congressional Record*, while "private accounts," "privatization plan," and four other variants show up among the most Democratic phrases.

Similarly, the large number of phrases relating to tax policy also accord well with expectations. We identify "death tax" (whose partisan pedigree we discuss above) as the third most Republican phrase. We identify "tax relief"—a term also advocated by Luntz (2005)—as strongly Republican, while "tax break" and "tax cuts for the wealthy" are strongly Democratic. Other phrases highlight the traditional partisan divide over the size of government—the Republican list includes four variants on "tax increase," while the Democratic list includes *sixteen* phrases referring to spending cuts ("cut student loans," "cut food stamps," "cut medicaid," and so forth).

On foreign policy, we identify variants on the phrase "global war on terror" as among the most strongly Republican phrases, while "war in Iraq" and "Iraq war" are Democratic. Stevenson (2005) describes the Bush administration's choice to adopt the phrase "global war on terror" to describe the conflict in the Middle East rather than explicitly referring to Iraq. Democratic phrases also include "veterans health care" and "bring our troops home"; Republican phrases include "Saddam Hussein," "change hearts and minds," and "Iraqi people."

that also appear on the list. Our results are robust to estimating our slant measure excluding these cases. Note that these phrases are not necessarily redundant statistically, because our preprocessing step (removal of stopwords and destemming) means that the two-word phrases may arise in somewhat different contexts than the three-word phrases.

### 3.2 Mapping Phrases to Ideology

We will measure ideology for each congressperson $c$ by the share $y_c$ of voters in the congressperson's constituency voting for the Republican presidential candidate in 2004. For each congressperson $c$, and phrase $p$, we also have data for each phrase on the share $s_{pc}$, defined as the frequency with which the congressperson uses phrase $p$, normalized as a share of the congressperson's total number of uses of the overall set of $1,000$ phrases.

We adopt a simple factor model of the relationship between language and ideology. We assume that the share of a phrase in a congressperson's speech ($s_{pc}$) is a linear function of her ideology ($y_c$):

$$s_{pc} = \alpha_p + \beta_p y_c + \varepsilon_{pc}, \tag{2}$$

where $\varepsilon_{pc}$ is an error term orthogonal to $y_c$. For notational ease, we will let $\tilde{s}_{pc} = s_{pc} - \alpha_p$ be the "de-meaned" frequency of phrase $p$ for congressperson $c$.

We will choose an estimator for $y_n$ to minimize a least-squares loss function, which penalizes an estimate $\bar{y}_n$ according to the Euclidean distance between the expected frequencies of each phrase and the observed (de-meaned) frequencies $\tilde{s}_{pn}$ for individual $n$. That is, we will choose $\bar{y}_n$ to solve

$$\bar{y}_n = \arg\min_{y_n} \sum_p \left(\tilde{s}_{pn} - \beta_p y_n\right)^2. \tag{3}$$

This problem is concave and has a unique, closed-form solution determined by its first-order condition:

$$\bar{y}_n = \frac{\sum_p \beta_p \tilde{s}_{pn}}{\sum_p \beta_p^2}. \tag{4}$$

This estimator is interpretable as a weighted average of the phrase frequencies $\tilde{s}_{pn}$, where the weights depend on the relationship between the the ideology of a congressperson and the frequency with which that congressperson uses the phrase.[19] If the use of some phrase $p$ is uncorrelated with a congressperson's ideology ($\beta_p = 0$), the use of that phrase does not contribute to the estimator

[19] The estimator also has a precedent in the text categorization literature: it is closely related to the "$K$-nearest neighbor" methodology, which would estimate $y_n$ by computing the average ideology $y_c$ of the $K$ congresspeople whose phrase frequencies are closest to person $n$ in terms of the Euclidean distance metric in equation (3). Our estimator takes a more parametric approach, which takes advantage of the continuous nature of the underlying ideology variable $y_c$.

$\bar{y}_n$. If phrase $p$ is used more often by more right-wing congresspeople ($\beta_p > 0$), the estimator will judge a person who uses $p$ often as more right-wing.

It is easy to see that $E(\bar{y}_n) = y_n$ if the shares $\tilde{s}_{pn}$ are governed by model (2). Note that the parameters $\beta_p$ are not observed by the econometrician, so that $\bar{y}_n$ cannot be computed directly. However, given counts by congressperson $c$ for each phrase $p$, it is straightforward to compute regression estimates $\hat{\alpha}_p, \hat{\beta}_p$ of the model's parameters, and then produce a (consistent) estimate $\hat{y}_n$ of $y_n$ by substituting these estimates for the true parameters $\alpha_p, \beta_p$ in equation (4).

The estimator $\hat{y}_n$ performs well in our sample of congresspeople. As we would expect based on its construction, a regression of estimated ideology $\hat{y}_c$ on true ideology $y_c$ across congresspeople produces a constant of 0 and a coefficient of 1, indicating that our estimator is a noisy but unbiased proxy for true ideology. Moreover, our estimator has a correlation of over 0.6 with true ideology, and a similarly high correlation with voting behavior (as measured by adjusted ADA scores).[20] These findings lend support to the expectation that our estimator will uncover genuine variation in ideological slant among newspapers.

### 3.3 Estimating Newspaper Slant

Our approach to measuring newspaper slant will be to treat each newspaper $n$ as an unknown congressperson as in the previous subsection, and to calculate the estimator $\hat{y}_n$ for each newspaper. This estimator answers the question, if this newspaper were a congressperson, how Republican would that congressperson's district be? To compute $\hat{y}_n$, we will use newspaper-level shares $s_{pn}$ of each of our final $1,000$ phrases. Recall that these counts are based only on a paper's news content; our search procedure excludes editorial and opinion pieces.[21] Our search also excludes national wire-service stories.

Despite these exclusions, our $1,000$ phrases are used an average of almost $14,000$ times in the content of papers in our sample in 2005. Even among newspapers in the bottom quartile of daily

[20] When we use adjusted ADA scores rather than vote shares to measure ideology $y_c$, the resulting estimator is highly correlated with our baseline measure, and produces very similar results in the analyses of newspaper slant we report below. See appendix A for details.

[21] We focus on news content both because policy discussions frequently focus on slant in the news, and because evidence suggests that readers pay significant attention to news content, even from local papers. Nearly two-thirds of Americans report getting news several times a week or daily from local newspapers (Harris Interactive 2006). Independent evidence suggests that almost 90 percent of readers of daily newspapers read the main news section, with over 80 percent reading the local news section (Newspaper Association of America 2006).

circulation in our sample, these phrases are used an average of nearly 4,000 times. One common reason for phrases associated with national issues to appear in the original content of local papers is coverage of local analogues of national issues. For example, although the repeal of the inheritance tax was initially a federal issue, a number of states considered similar repeals, prompting local papers to use language drawn from the federal debate. A headline from the November 21, 2005 issue of Salt Lake City's *Deseret News* spoke of "states not going along with federal repeal of *death-tax*" [italics added]. Around the same time, the *Hartford Courant* ran a front-page article proclaiming that "*Estate taxes* irk the rich; state may ease up but no repeal seen" [italics added]. Other occasions for local papers to use the phrases on our list include discussions of local impact of federal legislation, and of the actions of legislators from local districts.[22] On the whole, an informal examination of daily newspaper content suggests considerable scope for variation in the use of partisan language in these papers.

Across the newspapers in our sample, our slant measure correlates well with reader sentiment about the political leanings of different newspapers. For example, figure 1 shows a graph of our measure of slant for large papers against ratings of political orientation submitted by users to the media directory website Mondo Times (<http://www.mondotimes.com>).[23] The graph shows a clear association in the expected direction: papers rated as more conservative by Mondo Times users are also more Republican-leaning according to our index. Formal statistical tests confirm the visual evidence in figure 1. Across the 101 papers in our sample rated by more than one individual on the Mondo Times website, there is a correlation of 0.20 with our slant index ($p-value = 0.044$), and a rank correlation of 0.25 ($p - value = 0.011$). Note that we would not necessarily expect these correlations to be perfect, both because most papers receive only a few ratings, and because Mondo Times users are rating the editorial as well as news content of the papers, whereas our slant measure focuses on news content.

Our measure is also broadly in agreement with Groseclose and Milyo's (2005) bias measures for

[22] Note, however, that direct quotes of local congresspeople—which could cause a mechanical correlation between slant and the political leanings of local markets—comprise only a tiny fraction of the phrase mentions in our sample. Among 10 randomly chosen papers (representing different levels of circulation), we hand-coded the frequency of uses of the top 50 phrases in direct quotes of congresspeople. On average, such quotes account for only 0.3 percent of the phrase hits in this sample.

[23] We wish to thank Eric Kallgren of Mondo Code for graciously providing these data.

the six newspapers that were part of their analysis.[24] We find that the *New York Times*, *Los Angeles Times*, and *Washington Post* are similar to one another and to a fairly liberal congressperson; Groseclose and Milyo (2005) identify these papers as liberal and (statistically) fairly similar to one another. Groseclose and Milyo (2005) find that *USA Today* is somewhat closer to the center than these papers, a finding that we replicate but with a smaller magnitude. We also strongly confirm their finding that the *Washington Times* is significantly to the right of the other newspapers they consider. Our most significant point of disagreement is that we identify the *Wall Street Journal* as fairly right-leaning, whereas Groseclose and Milyo (2005) estimate that it is the most liberal newspaper in their sample.

# 4 The Demand for Slant

In this section, we study the relationship between newspaper slant and consumer demand for newspapers. We use zipcode-level data on newspaper circulation and political ideology to show, following Mullainathan and Shleifer (2005) and Gentzkow and Shapiro (2006), that households in more Republican zipcodes are more likely to read newspapers with a relatively right-wing slant. This evidence provides a useful check on the economic relevance of our slant measure, and allows us to compute, for each newspaper, the slant that would maximize its readership given the political ideology of consumers in its geographic market. We find that this "ideal slant" varies strongly with consumers' political beliefs, and that deviations from consumers' preferred slant involve a nontrivial sacrifice in circulation (and, hence, profits). These calculations serve as an important input to our study of the supply of newspaper slant in section 5.

## 4.1 Hotelling Model

We begin with a simple Hotelling model of newspaper demand. We denote the slant of newspaper $n$ by $y_n \in [0, 1]$. We assume that all households in zipcode $z$ have an ideal slant, which we will

[24] Groseclose and Milyo (2005) report results for six newspapers, nine television broadcasts, one radio broadcast, three national magazines, and one online news source.

model as a linear function of the zipcode's Republicanism $r_z$:

$$\tilde{y}_z = \alpha + \beta r_z \tag{5}$$

As discussed in the data section above, we will measure $r_z$ by the share of campaign contributions in the zipcode going to Republicans. The hypothesis that more conservative readers have a relatively greater taste for conservative newspapers implies that $\beta \geq 0$.

We assume that any difference between a newspaper's actual slant and a household's ideal slant imposes a quadratic loss (or "transport cost") on the household. Formally, we define the utility of household $i$, in zipcode $z$, for newspaper $n$ to be:

$$U_{izn} = -\gamma \left(y_n - \tilde{y}_z\right)^2 + \varepsilon_{zn} + \xi_{izn}. \tag{6}$$

Here $\varepsilon_{nz}$ is a zipcode-specific utility shock, $\xi_{izn}$ is a household-specific utility shock, and we expect $\gamma \geq 0$. Note the implicit restriction that all consumers within a given zipcode evaluate newspapers relative to the same ideal point. Though surely too strong, this assumption serves as a convenient approximation to a model in which the average Republican in a heavily Republican zipcode is further to the right than the average Republican in a more liberal zipcode.[25]

Following standard convention, we normalize the utility of the household's outside option—the consumption bundle that would be chosen conditional on not reading any newspaper—to 0. This incorporates an implicit maximization over all alternatives not written into the model, including television news, Internet news, and so forth. We also assume that the utility of consuming multiple newspapers is simply the sum of the newspapers' individual $U_{izn}$. This implies that a household reads a given newspaper $n$ if and only if $U_{izn} \geq 0$. This model imposes the assumption that different newspapers are independent in demand, and is thus a special case of a more general model where newspapers may be less-than-perfect substitutes. Evidence in Gentzkow (forthcoming) suggests that independence may be a reasonable approximation, and is likely to be closer to reality than

[25] We have estimated an alternative model in which each consumer's ideal slant depends on a household-specific ideology, drawn from a normal distribution whose mean varies across zipcodes. Retaining our other assumptions, we obtain a simple expression for consumer demand, and our findings regarding the determinants of newspaper demand are similar to the case in which we do not explicitly model within-zipcode heterogeneity in political ideology.

a standard discrete-choice framework that would require papers to be perfect substitutes at the individual level.[26]

Finally, we assume that the household-specific utility shock $\xi_{izn}$ is distributed i.i.d. uniform across households on the interval that includes the maximum and minimum values of $-\gamma\left(y_n - \tilde{y}_z\right)^2 + \varepsilon_{zn}$. This implies that the share of households reading newspaper $n$ in zipcode $z$ will be a linear function:

$$S_{zn} = \delta - \gamma\left(y_n - \tilde{y}_z\right)^2 + \varepsilon_{zn}, \tag{7}$$

where $\delta$ is a constant. (We abuse notation slightly here, in that both $\gamma$ and the variance of the zipcode-level shock $\varepsilon_{zn}$ are rescaled when we integrate over the household-specific shocks.) The assumption of uniform disturbances will simplify the analysis by making the interpretation of coefficients and the process of aggregation over zipcodes transparent. It is not critical, however—we have estimated an alternative model under the assumption that $\xi_{izn}$ is distributed i.i.d. type-II extreme value and obtain similar results.[27]

### 4.2 Identification and Estimation

There are two related sources of variation one could in principle use to identify the parameters of this model. One possibility would be to look at zipcodes with similar ideology and ask how the circulation of newspapers varies according to their slant. For any ideology $r_z$, it would be straightforward in principle to identify both the ideal point $\tilde{y}_z$ and the extent to which circulation falls when $y_n$ differs from $\tilde{y}_z$. Alternatively, one could look at newspapers with similar slant and compare circulation across zipcodes with different ideologies $r_z$. Because of the symmetry of the quadratic function, either or both of these sources of variation could allow an econometrician to recover the utility parameters $\alpha$, $\beta$, and $\gamma$.

This can be seen explicitly by substituting for $\tilde{y}_z$ in equation (7) and expanding the quadratic

[26] We have estimated a logit choice model in which each household is required to choose at most one of the newspapers available in its market. Our findings regarding consumer demand, and the implications for firms' profit-maximizing choice of slant, remain similar in this alternative specification. We have also experimented with excluding newspapers in multi-paper cities from our analysis, and find no meaningful change in results. See appendix A for details.

[27] See appendix A for details on this alternative specification.

to yield

$$S_{zn} = \left(\delta - \gamma\alpha^2\right) - \gamma\left[y_n^2 - 2\alpha y_n - 2\beta y_n r_z + 2\alpha\beta r_z + \beta^2 r_z^2\right] + \varepsilon_{zn} \tag{8}$$

The regression of $S_{zn}$ on a constant term and the linear, squared, and interaction terms in $y_n$ and $r_z$ has six free parameters; the model, however, has only four parameters: $\alpha$, $\beta$, $\gamma$, and $\delta$. One approach to estimation is to include zipcode fixed effects and so use only the first source of variation (across newspapers for a given zipcode). Another is to include newspaper fixed effects and so use only the second source of variation (across zipcodes for a given newspaper).

We will take the second approach, exploiting variation across zipcodes and controlling for mean differences among newspapers nonparametrically. We do this for two reasons. First, newspapers have a number of important characteristics (news quality, reputation, layout, etc.) that are likely to affect demand $S_{zn}$, could be correlated with slant, and are difficult to measure. In contrast, the most important zipcode-level shifters, such as education and income, can easily be controlled for using Census demographics. Second, the fact that a single newspaper may circulate in many different geographic markets introduces difficult-to-measure variation in the geographic "fit" between newspapers and individual zipcodes. Because there are many zipcodes in each locality, we can control for this fit flexibly by allowing different fixed effects for each newspaper in each market $m$ (defined as described above as either the zipcode's PMSA or its county).

Because we will take our identification from variation across zipcodes, we need to control explicitly for zipcode-specific shifters of demand. We include Census demographics such as education and income that make some zipcodes more prone to read newspapers than others. Moreover, we can model several dimensions of the fit between a zipcode and a newspaper by including interactions between zipcode demographics and the average level of the corresponding demographics in the newspaper's market. These controls will be important if non-political dimensions of fit are correlated with the political dimension we measure. For example, George and Waldfogel (2003) provide evidence that black consumers are more likely to read newspapers when the share of blacks in the overall market is large—presumably because newspapers react to this by shifting content in a way that appeals to these consumers. If blacks tend to be liberal, and if liberal content as captured in our slant measure is correlated with non-political content that appeals to blacks, this could cause us to overstate the magnitude of the coefficient on the interaction term $y_n r_z$. Our

controls will capture this kind of fit along a variety of demographic dimensions.

We thus estimate the following model:

$$S_{zn} = \delta_{mn} + 2\gamma\beta y_n r_z - 2\gamma\alpha\beta r_z - \gamma\beta^2 r_z^2 + X_z\phi_1 + W_{zn}\phi_2 + \varepsilon_{zn}. \tag{9}$$

Here, $\delta_{mn}$ are market-newspaper fixed effects, $X_z$ is a vector of observable zipcode characteristics, $W_{zn}$ is a vector of interactions between each characteristic of zipcode $z$ and the level of the same characteristic in the home market of newspaper $n$, and $\phi_1$ and $\phi_2$ are vectors of parameters.

A final econometric issue is that both slant ($y_n$) and zipcode Republicanism ($r_z$) are likely to be measured with error. In the case of $y_z$, the noise comes from the fact that our method for measuring slant is imperfect. Luckily, there is a natural instrument available for $y_n$: the overall share of Republicans in newspaper $n$'s market. If slant is correlated with the percent of the newspaper's market that is Republican (as our supply model below will predict), using this instrument amounts to asking whether newspapers from highly Republican markets have circulation that is relatively higher in Republican zipcodes within a given market. Formally, the instrument will be valid if the within-market correlation between zipcode Republicanism $r_z$ and the error term $\varepsilon_{zn}$ in the demand equation does not differ systematically with the share of Republicans in the newspaper's market. Since slant $y_n$ enters the regression only through the interaction $y_n r_z$, the instrument will be $R_n r_z$, where $R_n$ is market percent Republican. Note that instrumenting in this way will also correct for any endogeneity in slant $y_n$, though such a bias is made less likely by the fact that we are absorbing the main effect of $y_n$ in the fixed effects.[28]

In the case of $r_z$, noise is introduced because we are proxying for a zipcode's political tastes using the share of Republican campaign contributions, and the total number of contributions is

[28] An endogeneity concern would normally arise in a regression of demand on a product characteristic $y_n$ chosen optimally by firms, because market-level shocks to demand for newspapers would affect the firm's optimal choice of slant. However, the fixed effects $\delta_{mn}$ absorb any shocks to the taste for newspapers at the market level. A more subtle concern is that the distribution of demand shocks $\varepsilon_{zn}$ across zipcodes within a market affects the marginal return to slant—in particular, if in a market where Republican zipcodes randomly draw high shocks $\varepsilon_{zn}$ the optimal choice of slant shifts to the right. The resulting upward bias in the coefficient on $y_n r_z$ would then be corrected by our instrumental variables strategy under the same assumption stated in the text: that the within-market correlation between $r_z$ and $\varepsilon_{zn}$ does not differ systematically with the share voting Republican in the newspaper's home market. A final possibility—that zipcode ideology is itself affected by the slant of newspapers that circulate in the zipcode—requires the use of instruments that affect zipcode ideology but are plausibly unaffected by news content. A model in which we instrument for zipcode ideology with zipcode demographic characteristics produces estimates very similar to those of our baseline model.

sometimes small. Although we have no zipcode-level instruments for $r_z$, a different strategy is available because we can model explicitly the sampling variance in our measure of $r_z$, as a function of the total number of contributions received by either party. In particular, we can suppose that the share $r_z$ donating to Republican candidates is distributed binomially, with the probability of success given by $r_z^*$, the "true" ideology of the zipcode. Under this assumption, by standard arguments we can approximate $r_z$ by

$$r_z = r_z^* + \frac{r_z^*\left(1 - r_z^*\right)}{T_z}\upsilon_z \tag{10}$$

where $T_z$ is the total number of contributions in zipcode $z$ and $\upsilon_z$ is a standard normal disturbance. Although true ideology $r_z^*$ is by definition unknown, we can approximate the variance of the measurement error by assuming that $r_z^*$ is equal to its sample average. It is then straightforward to correct the coefficients using regression calibration (Fuller 1987).[29]

### 4.3 Results

Identification of equation (9) will be driven by the way the circulation of newspapers with similar slant $y_n$ varies across zipcodes with different ideologies $r_z$. The model makes two predictions about the form this variation should take. First, more Republican newspapers should circulate relatively more in Republican areas—the coefficient on the interaction $y_n r_z$ should be positive. This prediction is independent of the specific functional form we have chosen and would hold in a broad class of models. Second, there are interior ideal points—the share reading a paper with slant $y_n$ will take the form of an inverted U, highest in zipcodes where $r_z = \left(y_n - \alpha\right)/\beta$ and dropping off in zipcodes where $r_z$ is either higher or lower. This is a much finer prediction, dependent on second-order properties of the model.

Before turning to estimates of equation (9), we examine the extent to which these predictions

[29] In particular, we assume that the true ideology $r_z^*$ of each zipcode is distributed normally, with a mean that may depend on zipcode characteristics and a variance that we can estimate directly given an estimate of the variance of the sampling error. We estimate the (conditional) mean of the distribution of Republican shares by regressing shares for each zipcode on our full set of controls, including dummies for geographic market interacted with newspaper. We then compute, for each zipcode, the Bayesian posterior expectation of its true Republican share given our data on the share donating to Republicans. We repeat this exercise to compute the posterior expectation of the square of the true Republican share. We then estimate model (9) using these posterior expectations, rather than the observed share donating to Republicans, as independent variables.

are confirmed directly in the data. A relatively nonparametric way to look at the interaction between $y_n$ and $r_z$ is to estimate the coefficient on $r_z$ separately for each newspaper and ask how the coefficients vary with $y_n$. We have done this in a model analogous to equation (9). Figure 2 shows the estimated coefficients, plotted against our slant measure. The figure shows data for the 60 newspapers that circulate in markets containing more than 300 zipcodes, because these are the newspapers that provide the richest variation for identifying model (9). As predicted, the effect of zipcode Republicanism on circulation has a clear positive relationship with slant.

Figure 3 presents a different cut of the data that allows us to examine the stronger prediction of interior ideal points. Each panel shows the share reading newspapers in different deciles of the distribution of zipcode Republicanism $r_z$, after controlling for market-newspaper fixed effects $\delta_{mn}$ and weighted by the number of households in each zipcode. The first panel shows this relationship for newspapers in the lowest quartile of $y_n$, the second panel shows the relationship for papers in the second quartile, and so forth. Although far from perfectly clean, the graphs provide strong support for the existence of interior ideal points that shift to the right at higher levels of $y_n$.

Table 2 presents our estimates of equation (9). Column (1) shows the simplest OLS specification, in which we omit the controls $X_z$ and $W_{zn}$, do not instrument for slant $y_n$, and do not correct for measurement error in $r_z$. The results confirm a strong and statistically significant interaction between zipcode politics and newspaper slant, as suggested by figure 2. As predicted by the model, both the main effect of $r_z$ and its square enter negatively in determining demand.

The bottom rows of the table list the implied structural parameters from equation (9), $\alpha$, $\beta$, and $\gamma$, which can be computed through simple algebraic manipulation of the regression coefficients. We have also computed Monte Carlo confidence intervals for these parameters, by sampling from a multivariate normal distribution whose mean and variance-covariance matrix is given by the asymptotic covariance matrix of the estimated coefficients. All of these parameters have the expected sign and all except $\alpha$ are significantly different from zero.

In column (2) of the table, we correct for measurement error in $y_n$ by instrumenting for the interaction $y_n r_z$ using the interaction between the share of Republican voters in newspaper $n$'s market $R_n$ and $r_z$. As predicted, this causes the magnitudes of the coefficients to increase, but

the qualitative pattern to remain unchanged.[30] It also increases the precision of the structural parameters, $\alpha$ and $\beta$, that govern the relationship between a zipcode's ideology and its preferred newspaper slant. In column (3) of the table, we correct for measurement error in $r_z$ using regression calibration as discussed above. This causes the magnitudes of the coefficients to increase, and results in further improvements in the precision of the structural parameters.

The next two columns add controls for observable characteristics to the regression. Column (4) adds controls for zipcode demographics $X_z$. Column (5) includes these controls as well as interactions between zipcode demographics and analogous characteristics $W_{zn}$ measured at the level of the newspaper's markets. Neither set of controls substantially changes the estimated importance of the match between zipcode ideology and newspaper slant. Further specification checks, using a model in which we allow each zipcode to have its own taste for news, confirm a strong and robust interaction between zipcode ideology and newspaper slant.[31]

Our final estimates of the structural parameters in column (5) indicate that a zipcode in which all political contributions go to Democrats prefers a newspaper with slant 0.40, and that the ideal slant moves by a statistically significant 0.009 with every 10 percentage point change in the share contributing to Republicans in the zipcode. The positive and statistically significant estimate of $\gamma$ implies that deviations from a zipcode's ideal slant do indeed result in a loss of utility. To get a sense for the magnitude of the effect, note that the standard deviation of our slant measure is approximately 0.04, which is about two-thirds of the difference between the *Washington Post* and the *Washington Times*. Shifting a paper from a zipcode's ideal point (where $y_n = \tilde{y}_n$) to a level of slant one standard deviation away (where $y_n - \tilde{y}_n = 0.04$) would reduce the fraction of households reading by 3.4 percent, equivalent to a reduction of 9 percentage points in the share of households in the zipcode with a college education.

[30] The first stage regression underlying the 2SLS model shows a large and highly statistically significant effect of $R_n r_z$ on $y_n r_z$, ruling out any significant weak instruments concerns (Stock and Yogo 2002).

[31] More precisely, we have estimated a specification that includes *both* market-newspaper and zipcode fixed effects. Note that in this case it is impossible to identify all of the structural parameters, as the terms in $r_z$ and $r_z^2$ are absorbed in the fixed effects. Nevertheless, the coefficient $2\gamma\beta$ on the interaction term $y_n r_z$ *is* identified, and we find that it is large and statistically significantly positive. This provides strong evidence in favor of the prediction that more conservative consumers have a relatively higher taste for conservative news.

### 4.4 Computing the Profit-Maximizing Choice of Slant

Our estimates of model (9) allow us to calculate the slant that would be chosen by a newspaper that independently maximized its own profit. We write the variable profit of newspaper $n$ as:

$$\Pi_n = m_n \sum_z H_z S_{zn} \tag{11}$$

where $m_n$ is $n$'s average per-household markup (including both circulation and advertising revenue), $H_z$ is the number of households in $z$, and the sum is taken over all zipcodes in $n$'s primary market.[32] Note that we assume for simplicity that the per-household markup does not vary across zipcodes. The estimated ideal points are virtually identical when we relax this assumption and allow the value of consumers to advertisers to vary according to zipcode demographics.[33]

Substituting for $S_{zn}$ using equation (7) and maximizing with respect to $y_n$ yields the first-order condition:

$$\sum_z H_z \left(y_n - \alpha - \beta r_z\right) = 0. \tag{12}$$

Solving for $y_n$, we obtain the profit-maximizing level of slant, which we denote $ideal_n$:

$$ideal_n = \alpha + \beta \bar{r}_n \tag{13}$$

where $\bar{r}_n$ is the average share donating to Republican candidates in the market as a whole. Note that equation (13) is simply the expression for an individual zipcode's ideal point, with the zipcode level contribution share $r_z$ replaced by the market average $\bar{r}_n$. To eliminate any remaining measurement error in the market-level average share donating to Republicans, we regress this average on the share of voters $R_n$ voting for George W. Bush in the 2004 presidential election, and use predicted

[32] We restrict attention to a newspaper's primary market both for simplicity and because we do not have detailed circulation data for newspapers not covered in the ABC data. For the subset of newspapers for which it is possible to compute consumer ideal points using all markets in which the newspaper circulates, use of this more expansive market definition produces results similar to those we report below. While we have computed our slant measure for a total of 417 papers, we exclude four papers—the *New York Times*, the *Wall Street Journal*, the *Christian Science Monitor*, and *USA Today*—from our analysis of consumer ideal points because these national papers do not serve a well-defined local market. These exclusions do not meaningfully affect our results.

[33] We perform this calculation by estimating a model in which we regress each newspaper's reported ad rates (taken from the E&P database) on the demographic characteristics of its primary market. We then apply the coefficients from this regression to each zipcode's demographic characteristics, to produce an estimate of the relative advertising value of each zipcode. Finally, we adjust equation (11) to include zipcode-specific weights that depend on each zipcode's predicted advertising value.

values from this regression to calculate the ideal point $ideal_n$ for each newspaper.[34] In other words, our measure of $ideal_n$ will be a linear function of Bush's share of the two-party vote, parametrized by equation (13) and by the relationship between $\bar{r}_n$ and $R_n$.

# 5 The Supply of Slant

The previous section establishes that consumers are more likely to read a newspaper whose slant is close to their own political ideology. This creates a strong economic incentive for newspapers to tailor their slant to suit the political leanings of their geographic market, as measured by the newspaper-specific ideal points $ideal_n$ calculated above. In this section, we develop a model of the supply of slant that incorporates this incentive, along with non-pecuniary motives on the part of newspaper owners. We then examine the way our estimated ideal points compare with the actual slant chosen by newspapers, and test for the role of these additional incentives in determining slant.

## 5.1 Model

We assume that newspaper owners maximize an objective function that incorporates both profit and a direct concern for the ideological position of the newspaper. The latter may come from a variety of sources, including a desire to change the political views of readers (Balan, DeGraba, and Wickelgren 2005; Gentzkow, Glaeser, and Goldin 2006), direct utility from producing content that accords with the owner's own beliefs, and indirect incentives introduced by politicians (Besley and Prat 2006). Because our model combines both pecuniary and non-pecuniary motives, it is closely related to Becker's (1957) investigation of discrimination, in which employers may care about maximizing profits as well as about the identity of their employees. As in Becker's (1957) model, firms in our framework may face a trade-off between maximizing financial returns and satisfying their ideological tastes.

To capture non-pecuniary motives, we assume that each firm $g$ has an ideal slant $\mu_g$, to which

[34] Across markets, the share voting for Bush in 2004 and the share of donations going to Republican candidates from 2000 to 2004 have a correlation coefficient of nearly 0.8. Correlation patterns with respect to other proxies for market ideology generally show a weak relationship with donations, once vote shares are taken into account. These findings suggest that, while both measures are highly related statistically, vote shares are likely to be more precisely measured. This is not surprising given that the number of voters in a market is orders of magnitude larger than the number of donations to political candidates.

it would like its newspapers to conform. Although we will experiment with several direct measures of $\mu_g$, our primary approach will be to assume that owner ideology $\mu_g$ is a normally distributed random effect with mean $\bar{\mu}$ and variance $\sigma^2_\mu$. Paralleling our model of demand above, we assume that owners suffer a quadratic loss for each newspaper that deviates from their ideal point. We assume further that this loss is proportional to the number of households in the newspaper's market—a crude way to capture the intuition that a newspaper owner would obtain more private benefits from maintaining its preferred ideological position in a major city paper than in a small-town paper.[35]

We assume that firm $g$ chooses slant $y_n$ for each of its newspapers to maximize the following objective function:

$$V_g = \sum_n \left[ \Pi_n - \lambda H_n \left( y_n - \mu_g \right)^2 \right] \tag{14}$$

where $H_n$ is the number of households in newspaper $n$'s market. Recall that dollar profits $\Pi_n$ are simply the product of total demand for the newspaper (governed by the Hotelling model we estimate in section 4) and a per-reader markup. Substituting the demand model from equation (7), taking the derivative with respect to $y_n$, and solving yields an expression for the firm's optimal choice of slant $y^*_n$:

$$y^*_n = \frac{m_n \gamma}{m_n \gamma + \lambda} ideal_n + \frac{\lambda}{m_n \gamma + \lambda} \mu_g \tag{15}$$

where $m_n$ is the average markup of newspaper $n$'s consumers.

Equation (15) is straightforward to interpret: optimal slant is a weighted average of the profit-maximizing level of slant $ideal_n$ and the owner's preferred slant $\mu_g$, where the weights depend on the strength of the owner's tastes relative to the lost profits from deviations from consumer preferences. When the owner's tastes are strong relative to the dollar value of a marginal consumer (i.e., when $\lambda$ is large relative to $m_n$), slant will be close to the owners' preferred point $\mu_g$. By contrast, when consumers are valuable or the owner's tastes are weak (high $m_n$ or low $\lambda$), slant will be close to the profit-maximizing point $ideal_n$.

To implement equation (15) empirically, we suppose that our measure of slant $\hat{y}_n$ is equal to $y^*_n$

[35] If ideological benefits do not scale with the population of the geographic market, then the model will predict more conformity to consumers' tastes in larger markets than in smaller markets, as in Gentzkow, Glaeser, and Goldin (2006).

plus a noise term $\zeta_n$. Substituting for the mean of the owners' tastes $\bar{\mu}$ yields a standard random effects model of slant:

$$\hat{y}_n = \frac{\lambda}{m_n\gamma + \lambda}\bar{\mu} + \frac{m_n\gamma}{m_n\gamma + \lambda}ideal_n + \frac{\lambda}{m_n\gamma + \lambda}\tilde{\mu}_g + \zeta_n. \tag{16}$$

Here, the random effect term $\tilde{\mu}_g$ is distributed normally with mean 0 and variance $\sigma^2_\mu$.

Equation (16) implies that two features of the data can identify the economic relevance of owners' tastes. First, if ownership is important the residuals of a regression of $\hat{y}_n$ on $ideal_n$ should be correlated within ownership groups. Second, if the weight of tastes in the objective function $\lambda$ is large, $\hat{y}_n$ will be "pulled" toward owners' tastes, and will therefore move less than one-for-one with a change in $ideal_n$. In the limit case, as $\lambda$ tends to zero, measured slant is simply a noisy proxy for consumer preferences, so that a regression of $\hat{y}_n$ on $ideal_n$ will yield a constant of 0, a coefficient of 1, and residuals that are not correlated across papers with the same owner.

### 5.2 Reduced-Form Evidence

In this section, we compare our estimated ideal points with the actual slant chosen by newspapers, and present reduced-form evidence on the role of ownership. In the next section, we present estimates of equation (16) that allow us to assess the overall role of consumer characteristics and owners' tastes in determining equilibrium slant.

We first ask whether newspapers appear to deviate from the profit-maximizing level of slant on average. Average slant could deviate from the profit-maximizing level if it is influenced by reporters' preferences (Baron 2006), pressure from incumbent politicians (Besley and Prat 2006), or the average tastes of owners (Balan, DeGraba, and Wickelgren 2005). A large popular literature has argued that such forces create an overall liberal (Coulter 2003; Goldberg 2003) or conservative (Alterman 2003; Franken 2003) bias in the media. Our profit-maximizing model provides a well-defined benchmark against which to evaluate such claims.

We find no evidence of deviations on average from profit-maximization: the average level of slant in our sample of 413 newspapers is 0.47, while the average profit-maximizing point is 0.46. Figure 4 displays a histogram of the difference between actual slant and ideal point. Although the average

newspaper is slightly to the right of its consumers' preferences, this difference is economically small and statistically insignificant.[36]

While our results do not suggest systematic deviation from profit-maximization on average, they are nevertheless consistent with Groseclose and Milyo's (2005) finding that average news content resembles a left-of-center congressperson. We estimate that the average newspaper's language is most similar to that of a congressperson from a 47 percent Republican district, while in the average newspaper's market 53 percent of votes went to Bush in 2004.[37] However, we find that the average profit-maximizing point is *also* to the left of the average congressperson, and considerably closer to the average level of slant we observe. One possible explanation is that the consumers with the highest propensity to read—or whose readership is most sensitive to slant—tend to be to the left of the median voter.

We turn next to the question of whether variation across markets in the preferences of consumers can explain variation in newspaper slant. In figure 5, we graph the slant of a newspaper against the percent Republican in the newspaper's market, and plot a line showing our estimate of the ideal points $ideal_n$. Recall from section 4.4 that $ideal_n$ varies across news markets as a linear function of the share Republican, so the ideal points appear as a straight line. The graph shows clearly that in more Republican markets, newspapers adopt a more right-wing slant, exactly as predicted by the cross-market variation in consumer ideal points. The correlation between the actual and profit-maximizing levels of slant is 0.44, which is highly statistically significant ($p < 0.001$). Put differently, variation in consumer preferences explains nearly 20 percent of the variation in slant.

Recall that equation (16) implies that, if owners are willing to pay to tilt the news toward their own ideology, measured slant should move less than one-for-one with profit-maximizing slant. Comparing the estimated ideal points with the data suggests that slant actually varies *more* than one-for-one with $ideal_n$. Figure 5 thus provides some early evidence that ownership may be relatively unimportant. However, the slope of $ideal_n$ depends on the structural parameter $\beta$ from our

[36] Note that the appropriate notion of statistical significance here is one that incorporates both sampling uncertainty in the mean level of slant, and the uncertainty in the demand parameters that underly our calculation of the profit-maximizing level of slant. The confidence interval displayed in figure 4 takes account of both sources of uncertainty.

[37] Indeed, 68 percent of the newspapers in our sample have a slant below the share voting Republican in their primary markets, and a $t$-test definitively rejects the null hypothesis that the distribution of slant has the same mean as the distribution of Republican vote shares ($p < 0.001$). Economically, the difference between slant and Republican vote shares is large, representing almost 1.5 standard deviations of our slant measure.

demand model, and uncertainty in the estimate of $\beta$ must be incorporated into inferences about the relationship between $ideal_n$ and actual slant. After taking account of this uncertainty, neither a one-for-one relationship nor a less-than-one-for-one relationship can be rejected statistically.

A more direct (and statistically more powerful) way to look at the importance of ownership is to exploit the fact that many newspapers in our sample share the same owner. In the next subsection, we use this variation to identify the random-effect component of equation (16). First, however, we look at reduced-form evidence on the extent to which slant is correlated across newspapers with the same owner.

These results are presented in table 3. As a first look at the patterns in the data, column (1) presents the coefficient from a regression of each newspaper's slant on the average slant of other newspapers with the same owner. We find a statistically significant relationship between these two variables. This could arise either because owners influence slant, or because newspaper groups have a strong tendency to be clustered geographically, with owners specializing in owning regional, or even local, groups of newspapers (Lacy and Simon 1997; Martin 2003). Because political preferences have a strong geographic component (Fiorina 2005; Glaeser and Ward 2006), this clustering could create a spurious correlation between the slant of a newspaper and that of co-owned papers.

The remaining columns of table 3 attempt to separate these stories. Column (2) re-estimates the model of column (1), adjusting for each paper's profit-maximizing slant $ideal_n$. Specifically, we regress slant on $ideal_n$, and extract the residuals from this regression. We then regress each paper's own residuals on the average residuals of other newspapers with the same owner. This correction reduces the estimated relationship between a newspaper's slant and the slant of co-owned papers by about 10 percent. To correct more directly for geographic clustering, column (3) of table 3 includes a set of Census division dummies in the specification. Inclusion of these controls results in a much more muted estimate of the effect of ownership on slant than in the uncorrected estimates from column (1). Finally, in column (4), we include controls for state, and find a statistically insignificant effect of the slant of co-owned papers on a newspaper's slant, equivalent in magnitude to a change of 2.4 percentage points in a newspaper's slant for every 10 percentage point change in the average slant of co-owned papers. After adjusting for the allocation of ownership groups across states, we find no evidence of a statistical relationship between a newspaper's slant and the slant of co-owned

papers.

Figure 6 presents the findings in columns (1) and (4) graphically. In the first part of the figure, we show the unadjusted relationship between a newspaper's slant and that of co-owned papers. The positive relationship between these two variables suggested by column (1) of table 3 is visible in this graph. In the second part of this figure we adjust our measure of newspaper slant for the political attitudes and state of the newspaper's primary market. As the figure shows—and as column (4) of table 3 confirms—the adjusted relationship is essentially flat, indicating very little, if any, effect of ownership on the slant of a newspaper.

### 5.3 Random-Effects Estimates

The findings in table 3 suggest that owners exert a relatively small influence on newspaper slant, once the geographic clustering of ownership groups is taken into account. For a more quantitative evaluation of the importance of both ownership and consumer characteristics, table 4 presents estimates of the model in equation (16), allowing each owner to have an owner-specific random effect on the slant of its newspapers. We estimate the model by maximum likelihood, identifying the variance of the owner-specific random effect using information on the covariance between a newspaper's slant and that of co-owned papers. In other words, the random effects estimates translate the correlations highlighted in table 3 into a model of the underlying sources of variance in newspaper slant.

Column (1) of table 4 examines the extent to which slant is correlated within ownership groups before adjusting for the profit-maximizing level of slant. In a regression of slant on a constant term and the random effects $\tilde{\mu}_g$ alone, we estimate an ownership effect with a standard deviation of about 0.014. This standard deviation is statistically different from zero and economically nontrivial, accounting for 13 percent of the overall variation in measured slant.

Column (2) isolates the consumer side of the equation, controlling for ownership nonparametrically by including owner-specific fixed effects. There is substantial variation in Republican vote shares (and therefore in consumers' preferred slant) across markets within the same ownership group. For example, the markets where the *New York Times* company owns newspapers range from New York City (28 percent Republican), to Sarasota, FL (55 percent Republican), to Spar-

tanburg, SC (67 percent Republican). The effect of the profit-maximizing level of slant in this specification is strongly significant, suggesting that the correlation between slant and consumer characteristics in figure 5 was not an artifact of owners' clustering in politically similar markets. Consumer characteristics in this specification explain 19 percent of the remaining variation once the owner fixed effects are partialled out.

Column (3) presents estimates of equation (16), including both predicted ideal point and owner-specific random effects. When owner and consumer characteristics are combined, the share of variation explained by ownership falls from 13 percent to 9 percent, while the share explained by consumer characteristics remains roughly unchanged at 20 percent. This suggests that some of the ownership effect in specification (1) reflected correlation between the distribution of ownership groups and consumer characteristics rather than a causal effect of ownership.

In columns (4) and (5) of table 4, we add Census division and state fixed effects, respectively. If owners cluster geographically (Lacy and Simon 1997; Martin 2003), and newspapers in different areas use somewhat different language, this could bias upward our estimates of the effect of owners on newspaper slant. Consistent with the reduced-form results in table 3, column (4) shows that adding Census division controls further reduces estimated standard deviation of the ownership effect. In this case, we find that ownership explains a statistically insignificant 2 percent of the within-division slant, and our confidence intervals are precise enough to rule out ownership effects accounting for more than 6 percent of the explained variation in slant. Column (5) shows that adding fixed effects for the state in which the newspaper is located eliminates the estimated effect of ownership entirely, with confidence intervals that allow us to rule out effects larger than 12 percent of the within-state variation in slant. In contrast, the role of consumer characteristics actually grows stronger as we focus on variation in slant within geographic areas.

### 5.4 Structural Parameters and Interpretation

On the whole, the estimates of our model suggest that the variation in consumer characteristics captured by our estimated ideal points has a robust and economically important relationship with observed slant, consistently explaining roughly 20 percent of the variation in the sample. In contrast, the within-group correlation of slant appears to be largely an artifact of geographic clustering

of ownership groups. After controlling for the geographic clustering of owners and the political preferences of their consumers, we find that variation in ownership explains little or none of the variation in slant. Stating the result a different way, moving from the current level of cross-market heterogeneity in consumer preferences to a world in which all newspapers cater to markets with identical political preferences would reduce the diversity of slant in our preferred specification by 22 percent (with the top of the confidence interval at 30 percent), whereas moving to a world with a single newspaper owner would have a negligible effect (with the top of the confidence interval at 12 percent).

Recall that even among congresspeople, a significant portion of our slant measure is uncorrelated with true ideology. By assuming that the degree of measurement error among newspapers is similar to that among congresspeople, we can adjust these variance counterfactuals for the degree of measurement error in slant, and thus convert them to effects on "true" (as opposed to measured) slant. Among congresspeople, approximately 37 percent of the variance in our slant measure is related to true ideology, implying that, among congresspeople, about 63 percent of the variation in our slant index is measurement error. Assuming this share is applicable to newspapers, we can scale up the variance counterfactuals described above, expressing them in terms of true underlying slant rather than measured slant. This rescaling implies that eliminating cross-market heterogeneity in consumer preferences would eliminate fully 54 percent of the true variation in slant.

We turn next to an interpretation of our results in terms of the structural model of subsection 5.1. Equation 16 involves two key structural parameters: the variance of owner tastes $\sigma_\mu^2$, and the parameter $\lambda$ that governs how important these tastes are relative to profits. Table 4 reports the overall share of variance in slant attributable to ownership—a statistic that combines the effect of $\lambda$ and $\sigma_\mu^2$.

The point estimate in the final column of table 4 for the standard deviation of the ownership effect—that is, of $\frac{\lambda}{m_n\gamma+\lambda}\tilde{\mu}_g$ in equation 16—is zero. Assuming that owners do not all have identical ideal points $\mu$ ($\sigma_\mu^2 \neq 0$), this implies that $\lambda = 0$ and that any positive value of $\sigma_\mu^2$ is consistent with the data. We will therefore not focus on this point estimate, but on the structural parameters implied by the highest standard deviation of the ownership effect admitted by our 95 percent confidence interval (0.01). Note that the latter value is also approximately the point estimate in a

model that does not control for the geographic clustering of ownership groups (column (3) in table 4).

As already discussed, equation 16 suggests two features of the data that could identify the role of ownership: (i) the extent to which the residuals of slant on ideal point are correlated within ownership groups; and (ii) the extent to which the relationship between slant and ideal point is attenuated. In principle, these two sources of identification are sufficient to pin down $\lambda$ and $\sigma_\mu^2$ separately. However, as also discussed above, uncertainty in the first-stage demand estimates means that we have much more precision in estimating (i) than in estimating (ii). Rather than attempting to identify $\lambda$ and $\sigma_\mu^2$ directly, therefore, we will ask what value of $\lambda$ the model implies for several reasonable estimates of $\sigma_\mu^2$.

We ask, first, what the estimates would imply if owners were randomly drawn from the population of *zipcodes*—that is, if the variance of the owner ideal point $\mu$ were the same as the variance of consumer ideal points $\tilde{y}_z$ across zipcodes. This gives $\sigma_\mu^2 = 0.020$; an ownership effect at the top end of the 95 percent confidence interval would then imply $\lambda = 11$. One way to interpret these numbers is to recall that an owner's willingness to pay to broadcast her most preferred slant, $\mu_g$, rather than some alternative, $y_n$, is $\lambda \left(y_n - \mu_g\right)^2$ per reader per day. Given that $\lambda = 11$, this would imply that an owner would be willing to pay \$2.34 per reader per year to reduce the gap between actual and preferred slant by one standard deviation (approximately 0.04). Although this figure is not enormous, it is nontrivial. Thus, even if ideology varies significantly across owners, their willingness to pay to impose their preferences would have to be substantial in order to account for even a small amount of variation in slant.

To check the sensitivity of this conclusion, we note that if owners' tastes were randomly drawn from the population of *markets*, then $\sigma_\mu^2 = 0.016$, implying $\lambda = 20$. In this case, the typical owner would be willing to pay \$4.20 per reader per year to avoid a one-standard-deviation gap between the actual slant and her preferred slant. By contrast, if owners' preferences had an extreme distribution, in the sense that half of owners are like 100% Republican zipcodes and half are like 100% Democrat zipcodes, then $\sigma_\mu^2 = 0.045$, $\lambda = 3$, and the willingness to pay to avoid a one-standard-deviation gap between actual and preferred slant would be \$0.68 per reader per year. In all of these cases, accounting for the owner effects at the top of our confidence intervals would require a reasonably

substantial willingness to pay. This lends further economic credence to our finding that, in practice, owners exert a relatively small influence on newspaper slant.

### 5.5 Robustness

Having established a strong relationship between consumer characteristics and slant, we now check the robustness of this finding to a number of alternative specifications. Because we model consumer preferences as a linear function of the 2004 Republican vote share, these specification checks can be thought of as tests of the robustness of the relationship between vote shares and slant. The fact that we have scaled vote shares to measure the profit-maximizing level of slant $ideal_n$ gives an additional economic interpretation to the coefficients, but this interpretation must be taken with caution since we have not incorporated the specification checks and controls below into the demand model from which we predict $ideal_n$.

Our first two robustness checks address the possibility of reverse causality. Although we have been interpreting the coefficients on $ideal_n$ as reflecting an effect of consumer preferences on media slant, some of the relationship we estimate between consumer ideal points and observed slant could result from an effect of newspaper slant on voter beliefs and behavior (Gentzkow and Shapiro 2004; Della Vigna and Kaplan 2007; Gerber, Karlan, and Bergan 2006). In column (1) of table 5, we instrument for ideal slant with the share of DDB Needham survey respondents in the newspaper's market reporting that they attend church monthly or more during 1972-1998. This variable has a large effect on a market's political leaning (Glaeser, Ponzetto, and Shapiro 2005) but is unlikely to be a direct result of newspaper slant in 2005.[38] As the table shows, we continue to find a positive and statistically significant effect of consumer preferences on newspaper slant, with a coefficient that is similar to that observed in the regressions in table 4. In column (2) we instrument for slant with three other pre-determined characteristics of the newspaper's market: log population (to capture large urban markets), percent black, and percent with a college degree. All three are strong predictors of the share voting Republican, and when we instrument the coefficient on $ideal_n$ almost doubles. This provides additional evidence that the correlation we observe in the data is

[38] In a regression using data from the 406 news markets for which the church attendance variable is available, we find that an increase of 10 percentage points in the share of respondents attending church monthly is associated with an increase of about 0.2 percentage points in the profit-maximizing level of slant. This relationship is highly statistically significant ($p < 0.001$).

not driven by reverse-causality.

Although the implicit assumption in column (2) is that consumer demographics affect slant through their effect on political preferences, it is possible that a newspaper's use of language varies for non-political reasons with the characteristics of the local population, say because certain groups prefer certain words independent of their political connotations. In principle, such effects could confound our estimates of the impact of consumer preferences on slant. Note, however, that if these characteristics matter primarily through their impact on political attitudes, including them as controls will result in potentially misleading estimates of the true effect of consumer preferences on slant.

With this caveat in mind, column (3) adds log population, percent black, and percent with a college degree to the regression as controls. Our estimate of the effect of consumer preferences on newspaper slant decreases somewhat, but remains strong and statistically significant in this specification, and including a much wider vector of controls (specification not shown) results in similar estimates of our key coefficient. Note also that the direction of the coefficients on the demographic variables is consistent with their relationship to political preferences, indicating that these variables might be proxying for unmeasured dimensions of political heterogeneity (and hence of preferences for slant). The final column of table 5 includes the log of the newspaper's number of employees, the log of the number of pages, and the number of Pulitzer prizes from 1970-2000 as controls for newspaper quality (following Berry and Waldfogel 2003). While quality does appear to be negatively correlated with our measure of slant, including these controls, if anything, tends to increase somewhat the estimated effect of consumer preferences on slant.[39]

### 5.6 Changes in Consumer Preferences and Ownership

While we have focused primarily on cross-sectional tests of the effect of consumer preferences and owner identity on newspaper slant using 2005 data, we have also computed a preliminary version of our slant measure for the years 2000-2004. To compute this measure, we re-apply our procedure to the *Congressional Record* for each respective year, and search for the top 1,000 partisan phrases

[39] We have also computed a language-based index that predicts the sophistication of a congressperson's constituency (measured by the share of the constituency that is college-educated), given our set of partisan phrases. In appendix A, we show that our main results are robust to including this measure as a control, which supports the view that our measure captures partisanship, and not merely the sophistication of a newspaper's language.

using databases of news content for the relevant year. Because the availability of digital news content has risen over time, we have a larger sample of newspapers available in later years.

For our first test, we ask whether changes in consumer preferences are associated with changes in newspaper slant. Such a test could be confounded by reverse causality, but it may nevertheless be informative about the drivers of news content. To conduct a test of how newspapers respond to changes in political preferences, we have standardized the slant measure to have a mean of zero and a standard deviation of unity within each year. We then compute, for each of the newspapers available in both 2000 and 2004, the change in slant from 2000 to 2004, as well as the change in the share voting for Bush between the 2000 and 2004 presidential elections, among counties in the newspaper's market. We find that changes in slant and changes in vote shares are strongly correlated, and the relationship is both economically large (if anything larger than the relationship we estimate in the cross section) and statistically significant ($p < 0.001$).

A related question is whether changes in newspaper ownership are reflected in changes in their slant. We have identified three acquisitions during our sample period (2000-2005) for which we can measure slant annually for at least one paper owned by both the acquired and acquiring firm. First, in 2000, the Tribune Company (owner of the *Chicago Tribune*) acquired the Times-Mirror Corporation (owner of the *Los Angeles Times*). We have slant measures for seven papers owned by the combined company—5 initially owned by Times-Mirror and 2 initially owned by Tribune. Second, also in 2000, Thomson Corporation sold 21 daily papers to Gannett. Our sample includes 3 of these papers and 53 papers owned by Gannett prior to the change. Finally, in 2002, Lee Enterprises acquired all 16 dailies owned by Howard Newspapers. We have data on 1 of the Howard papers (the Twin Falls, ID *Times-News*) and on 6 papers previously owned by Lee.

A straightforward implication of the model in equation (16) is that if ownership effects are important, the difference between the mean slant of the acquired and acquiring papers should fall following the merger. Comparing the pre- and post-merger means shows that the difference did fall somewhat for the Times-Mirror-Tribune and Howard-Lee mergers, while it increased somewhat for the Thomson-Gannett merger. In none of these cases is the change statistically significant. This test has low power, due to the small number of mergers and the small number of papers involved in each. However, the results are consistent with our prior finding of a small or zero average effect

of ownership on slant.

# 6 Additional Determinants of Newspaper Slant

In this section, we enrich our model to allow for fixed costs in the production of news, as well as variation in the financial and political incentives to slant the news. We then test the predictions of the enriched models. We find little evidence that fixed costs cause a homogenization of news content. We also find no evidence that public companies or companies with larger per-reader markups display less owner-specific variation in slant. Finally, we show that direct proxies for owners' political attitudes are unrelated to slant, and that owners do not exert a greater influence in areas where the political returns to persuasion are highest. Taken together, these findings further support the view that owners exert at most a small or modest role on the ideological content of the news.

## 6.1 Fixed Costs in the Production of News

In section 5 above, we model owners' impact on newspaper slant by supposing that each owner has a preferred ideological position. It is also possible that owners have no inherent preference for one slant or another, but still like their papers to have similar slant so as to economize on firm-level fixed costs in the production of news content. There are several reasons this might be the case: if news stories can be written once and shared among multiple papers (and this sharing takes place more efficiently within a firm than between firms), if editorial staff for multiple papers receive training together, or if market research can be conducted jointly, for example. In any of these cases, it may be efficient to partially homogenize content across papers.

One simple way to model fixed costs is to assume that a firm $g$ pays a cost proportional to the heterogeneity in slant $y_n$ across the newspapers that it owns. Formally, we will measure heterogeneity by the sum of squared deviations in slant, so that each firm maximizes

$$V_g = \sum_n \Pi_n - \rho \sum_{n=1}^{N} H_n \left(y_n - \bar{y}_g\right)^2 \tag{17}$$

where $\bar{y}_g$ is the (household-weighted) mean slant among papers owned by firm $g$. It is straightfor-

ward to show that this model implies a simple linear form for optimal slant:

$$y_n^* = \frac{\gamma}{\gamma + \rho} ideal_j + \frac{\rho}{\gamma + \rho} \overline{ideal}_g \tag{18}$$

where $\overline{ideal}_g$ is the (household-weighted) average ideal point among papers owned by $g$ and the coefficients are functions of the underlying structural parameters. Intuitively, if fixed costs are important, all papers in a group should be pulled toward the slant that is optimal for the group's average market. Because this average ideal point is available from our demand estimates, we can test directly for the importance of fixed costs.

Figure 7 presents such a test, graphing a newspaper's slant against the average profit-maximizing point among co-owned papers, both partialled with respect to the newspaper's own profit-maximizing slant. We find only weak evidence that this variable influences newspaper slant, with a coefficient of 0.36 that is statistically insignificant ($p - value = 0.610$), and economically much smaller than the effect of the preferences of consumers in a newspaper's own market. The evidence in this figure suggests that a desire to homogenize the news to economize on production costs is not an important determinant of slant.

### 6.2 The Response of Slant to Financial Incentives

Equation (16) shows that, as the markup per consumer rises, newspaper slant will depend more on consumer preferences and less on owners' preferences, because the cost of deviating from consumer's preferences grows larger as the marginal reader becomes more valuable. For a simple test of this hypothesis, we have computed a dummy variable equal to one if a newspaper's per-reader advertising rate is above-median, and 0 otherwise. We can then estimate a version of equation (16) that allows the impact of both consumer preferences and owner ideology to depend on whether a newspaper has an above-median advertising rate.

Table 6 presents estimates of this model, showing the coefficient on the profit-maximizing level of slant, as well as the standard deviation of the owner random effect, for both above- and below-median advertising rate papers. Because we omit geographic controls from this table, the evidence we present here provides an independent test of whether the owner effects present in the model

without geographic controls are spurious. If these effects are indeed generated by owner ideology, then we would expect them to be larger in the low-advertising rate group. As the first two columns of the table show, this is not the case. If anything, owner effects are (statistically insignificantly) larger in the high-markup group, as is the effect of consumer preferences. This evidence therefore does not lend support to the view that owner ideology influences slant, and therefore tends to agree with our argument that these owner effects are most likely to be a spurious result of the endogenous geographic clustering of ownership groups.

A related hypothesis is that public firms—which may be constrained by capital markets to maximize profits without regard to ideology—cater more closely to the tastes of their customers, and display a smaller owner-specific effect on slant. The second two columns compute these effects separately for publicly and privately held firms. If anything, we find that publicly traded firms display greater owner effects and less sensitivity to consumer preferences, again casting doubt on the view that the estimated owner effects are due to a trade-off between financial and non-pecuniary interests.[40]

### 6.3 The Response of Slant to Political Incentives

In addition to studying the question of whether ownership is statistically related to the slant of a newspaper, we can also investigate whether slant is related to direct proxies for owners' political preferences. In figure 8, we show that newspaper slant is unrelated to the extent to which both corporate executives and the corporation itself contribute to Republican rather than Democratic campaigns. Although donations are by no means a perfect proxy for ideology, these findings cast further doubt on the importance of owner preferences in determining newspaper slant.

An alternative test of the role of political incentives is to ask, in the spirit of the tests in the previous subsection, whether slant responds more to owner identity in circumstances in which the political returns to changing consumers' voting behavior are large. To implement this test, we use our preliminary measure of slant in 2004, and test whether slant is more responsive to owner identity

[40] Interestingly, however, we find in unreported regressions that public firms display economically (though not statistically) significant "fixed cost" effects of the sort modeled in equation (18). In other words, newspapers in public firms tend to cater both to the tastes of consumers in their own markets, and to those of consumers in markets served by co-owned papers. This may result from larger firms' ability to achieve scale economies in the production of news, such as through firm-specific wire services.

(and less responsive to consumer preferences) in states that were considered to be "battleground" states during the 2004 presidential election. We find no evidence that this is the case, arguing against the view that owners shift the slant of their newspapers to achieve political aims.

Finally, as a test of Besley and Prat's (2006) hypothesis that the media will tend to favor the preferred policies of incumbents, we have estimated a regression that allows slant to vary with the party of the incumbent governor (as of the end of 2005), controlling for the preferences of consumers. We find that, controlling for the preferences of consumers, having a Republican governor is associated with a statistically insignificant reduction in slant of about 0.9 percentage points, with a confidence interval that rules out effects larger than about 0.4 percentage points (one-tenth of a standard deviation).

For a related test, we have computed the share of representatives to the U.S. House from districts in each newspaper's market who are Republican, as of the 109th Congress. Controlling for consumers' preferences, this share has a statistically insignificant negative effect on slant. Quantitatively, the coefficient is extremely small, indicating that moving from a completely Democratic to a completely Republican delegation reduces newspaper slant by 0.003, with a confidence interval that can rule out substantial positive effects. Though crude, these tests provide the first direct, large-scale empirical evidence on the impact of incumbent politicians on news content, and suggest that the party affiliation of incumbents does not significantly affect newspapers' political positioning.

### 6.4 Tastes of Reporters and Editors

The model estimated in section 5 allows for the possibility that newspaper owners have direct preferences over slant. A different possibility is that it is not the tastes of owners, but the tastes of reporters and editors that matter. For example, Baron (2006) develops a model in which workers are willing to accept lower wages to publish news slanted toward their personal views, and shows that this could affect slant in equilibrium. Importantly, if tastes of workers vary across local markets in a way similar to the tastes of consumers, this could induce correlation between actual slant and market-level politics, and possibly confound our estimates of the role of consumer preferences in driving slant.

Although we cannot estimate the importance of workers' tastes directly, we believe for two

reasons that they are unlikely to be an important confound to our findings. First, in order for market-level variation of reporters' and editors' tastes to matter, it must be that mobility across markets is limited—for example, because there is an economic advantage to newspapers of having reporters and editors drawn from the local population. Otherwise, newspapers would simply hire reporters and editors willing to adhere to the slant best suited to consumer demand. Several pieces of evidence suggest that newspapers are not confined to hiring local talent, and that, if anything, reporters and editors are more mobile than demographically similar professionals. According to one survey, the average college-educated journalist has nearly a 40 percent chance of working in a Census division other than the one in which she attended college (Weaver and Wilhoit 1996). This is considerably higher than the average among other college-educated workers, according to evidence from the 1979 National Longitudinal Study of Youth (NLSY).[41] Census data support the view that reporters are a highly mobile population, even compared with other highly educated professionals. Controlling for education, age, gender, and race, reporters and editors are 8 percentage points more likely to live in a state other than the one in which they were born.[42] Additionally, the labor market does not appear to assign any premium to local talent in the market for reporters and editors. Reporters and editors born outside their current state of residence earn, if anything, somewhat *more* than those working in their states of nativity. Although this effect may be due in part to unmeasured variation in human capital (Wozniak 2006), combined with the evidence on mobility patterns it provides little support for the view that newspaper owners find it economically advantageous to hire locally.

Second, our model allows us to calibrate the magnitude of tastes for slant that would be necessary to generate the variance in slant we observe in our data. Recall that our demand estimates imply that choosing slant one standard deviation from consumers' preferred level of slant would reduce circulation (and, hence, variable profits) by about 3.4 percent. If an average newspaper were to deviate by one standard deviation from the optimal slant because of reporters' tastes, it would have to be the case that hiring equally qualified reporters willing to produce at the optimal slant

[41] We are extremely grateful to Lisa Kahn for providing the appropriate calculations from the NLSY.

[42] They are also three percentage points more likely to have moved in the past five years. These figures are coefficients on reporter/editor dummies in regressions using data from the 1970, 1980, 1990, and 2000 Censuses (Ruggles et al, 2004). The sample is restricted to 25- to 55-year-old workers in professional occupations (1950 occupation codes 000-099). Wage regressions reported below are restricted to prime-age male reporters and editors working full-time.

would cost the firm *more* than 3.4 percent of variable profits. To get a sense of the wage effect this would imply, calculations based on data in Gentzkow (forthcoming) suggest that the *Washington Post*'s variable profit in 2004 was on the order of $500 million.[43] *Burrelle's/Luce Media Directory 2001* (Burrelle's Information Services 2001) lists 222 reporters and 175 editors working for the *Post*. If we assume that the average reporter's salary is $75,000 per year and the average editor's salary is $125,000 per year (probably an overestimate),[44] we estimate the *Post*'s wage bill for reporters and editors to be about $43 million per year. This implies that the paper would have to be unable to hire staff willing to produce at the optimal slant even if it were willing to increase wages by (500/43)*3.5 percent = 41 percent. Although these estimates are rough, and the *Washington Post* is not necessarily a representative paper, they suggest that both the magnitude of tastes for slant and the barriers to mobility would have to be extremely large to explain a significant fraction of the variation of slant in our sample.

# 7 Conclusions

In this paper, we develop and estimate a new measure of slant that compares the use of partisan language in newspapers with that of Democrats and Republicans in Congress. Our measure is computable with a minimum of subjective input, is related to readers' subjective ratings of newspaper slant, and is available for newspapers representing over 70 percent of the daily circulation in the United States.

Combining our measure with zipcode-level circulation data, we show that consumer demand responds strongly to the fit between a newspaper's slant and the ideology of potential readers, implying an economic incentive for newspapers to tailor their slant to the ideological predispositions of consumers. We document such an effect, and show that variation in consumer preferences accounts for nearly one-fifth of the variation in measured slant in our sample.

By contrast, we find much less evidence for a role of newspaper owners in determining slant.

[43] Gentzkow (forthcoming) estimates that the variable profit per *daily* copy sold is $1.83. Applying the same profit rate to Sunday copies (probably an understatement) gives a total yearly variable profit of $539 million.

[44] The website <www.salary.com> provides estimates of the median salary and benefits by job title "based on broad national data, reported exclusively by human resource departments of tens of thousands of employers." The website's estimate of median total compensation for a mid-career reporter in the Washington DC market is $37,513. The median total compensation for a relatively junior editor is $56,232, for a supervising editor is $74,515, and for a managing editor is $84,727.

While slant is somewhat correlated across co-owned papers, this effect seems largely to be driven by the geographic clustering of ownership groups. After controlling for the geographic location of newspapers, we find no evidence that the variation in slant has an owner-specific component.

Taken together, our findings suggest that ownership diversity may not be a critical precondition for ideological diversity in the media. This conclusion has broad implications for the regulation of ownership in the media. While the use of partisan language is only one of many possible ways in which ownership might affect media content, our analysis is an important step in the direction of large-scale, quantitative studies of the determinants of news content. Beyond its methodological contribution to the measurement of slant, our work shows that consumers play a fundamental role in determining the ideological positioning of media outlets, suggesting that a product-differentiation framework may help to shed light on a wide range of public policy questions surrounding the news media, and information providers more generally.

# A Additional Robustness Checks

In this section, we discuss a number of checks on our main results; namely, that consumer preferences drive an important part of the variation in newspaper slant, and that ownership explains a much smaller share of this variation. We present the results of these alternative specifications in appendix table 1. In each row, we present the results of a random effects model of the form used in table 4, in which we regress a newspaper's slant on the preferred slant of its consumers, state fixed effects, and owner-specific random effects. The table presents the estimated effect of consumer preferences, and estimates of the share of the residual (within-state) variance attributable to variance in consumer preferences and ownership. For comparison, row (1) presents results from the final specification in table 4.

In row (2) of appendix table 1, we present our first robustness check, in which we re-estimate the demand model of section 4 using a logit (as opposed to linear) demand specification. That is, we assume that the household-specific choice error ($\xi_{izn}$ in equation 6) is distributed as type-II extreme value rather than uniformly. The result is that the shares $S_{zn}$ in equation (9) are replaced by the term $\log\left(S_{zn}/\left(1-S_{zn}\right)\right)$. After estimating this model, we compute numerically, for each newspaper, the slant that would maximize demand in the paper's home (headquarters) market. To eliminate measurement error in this calculation coming from noise in our zipcode-level ideology measure, we estimate a predicted value of the logit-demand-maximizing slant by regressing it on the share of votes in the newspaper's market going to George Bush in the 2004 presidential election. As the estimates in row (2) show, using this alternative functional form for demand does not affect our estimates of the importance of consumers and owners in determining cross-newspaper variation in slant.

In row (3) of appendix table 1, we present results using a logit demand system that explicitly incorporates substitution across newspapers. As in the previous specification, this model assumes that the household-specific choice error $\xi_{izn}$ is distributed type-II-extreme-value. However, rather than assuming that the utility of each newspaper does not depend on which other papers the household reads, it assumes that each household can choose at most one of the papers available in its market. The result is that the left-hand side of equation 9 becomes $\log\left(S_{zn}/S_{z0}\right)$, where $S_{z0}$ is the share of households reading no newspaper. After estimating this model, we compute for each newspaper the demand-maximizing level of slant given its market (and the slant of other newspapers available in the market), and predict this value using variation in the share voting for Bush in 2004. The estimates in row (3) show that this alternative demand system does not produce different conclusions regarding the role of consumer and owner heterogeneity in driving newspaper slant.

In row (4) of appendix table 1, we conduct a related robustness check, in which we exclude newspapers headquartered in multi-paper cities from our analysis. Excluding these papers provides an additional check on whether the fact that we have not modeled competition directly is a source of bias in our estimates. As the table shows, this exclusion does not meaningfully change our results.

In row (5) of appendix table 1, we show that our results are robust to controlling for a measure of the "sophistication" of the newspaper's language. Variation in sophistication could confound our estimates if, for example, more liberal markets tend also to be more educated, and hence prefer more sophisticated language. To measure the sophistication of a newspaper's language, we have estimated a version of our "slant" measure, but replacing congressperson ideology with a measure of the share of adults with a college degree or higher in the congressperson's constituency. In other words, the sophistication measure tells us how educated we would expect a congressperson's constituency to be given data on her use of our 1,000 partisan phrases. Similarly, the measure allows us to assess how educated a newspaper's constituency would be, if it were in Congress. As

row (4) shows, controlling for this variable does not meaningfully affect our results.

In row (6) of appendix table 1, we consider the robustness of our results to an alteration in our selection of partisan phrases. In particular, we tighten the cutoffs on the number of hits a phrase must have in newspaper headlines from 2000-2005 by setting them equal to the 5th and 95th percentiles in our sample. We then select from the remaining phrases the top 1,000 by $\chi^2$. Computing our slant measure based on this list results in essentially identical statistical results. This finding suggests that the minimum and maximum hit cutoffs we have imposed for computational efficiency are not a likely source of bias.

In row (7) of appendix table 1, we show results using an alternative measure of slant, generated by measuring a congressperson's ideology using her adjusted ADA score (Groseclose, Levitt, and Snyder 1999), rather than the presidential votes of her constituency. The adjusted ADA score measures the left-right orientation of a congressperson's roll call votes. Using this alternative measure does not yield different conclusions.

## References


[1] Kjersti Aas and Line Eikvil. Text categorisation: A survey. *Norwegian Computing Center Mimeograph*, June 1999.

[2] Roya Akhavan-Majid, Anita Rife, and Sheila Gopinath. Chain ownership and editorial independence: A case study of Gannett newspapers. *Journalism and Mass Communication Quarterly*, 68:59–66, Spring-Summer 1991.

[3] Eric Alterman. *What Liberal Media? The Truth about Bias and the News.* Basic Books, February 2003.

[4] Stephen Ansolabehere, Rebecca Lessem, and James M. Snyder Jr. The orientation of newspaper endorsements in U.S. elections, 1940-2002. *Quarterly Journal of Political Science*, 1(4):393–404, 2006.

[5] Werner Antweiler and Murray Z. Frank. Is all that talk just noise? The information content of Internet message boards. *Journal of Finance*, 59(3):1259–1294, June 2004.

[6] Ben H. Bagdikian. *The Media Monopoly.* Beacon Press, Boston, 6th edition, 2000.

[7] David J. Balan, Patrick DeGraba, and Abraham L. Wickelgren. Ideological persuasion in the media. *Federal Trade Commission mimeograph*, April 2005.

[8] David P. Baron. Persistent media bias. *Journal of Public Economics*, 90:1–36, 2006.

[9] Gary S. Becker. *The Economics of Discrimination.* University of Chicago Press, 2nd (1971) edition, 1957.

[10] Steven Berry and Joel Waldfogel. Product quality and market size. *NBER Working Paper No. 9675*, 2003.

[11] Timothy Besley and Andrea Prat. Handcuffs for the grabbing hand? Media capture and government accountability. *American Economic Review*, 96(3):720–736, June 2006.

[12] Burrelle's Information Services. *Burrelle's/Luce Media Directory - 2001 Edition.* Livingston, NJ, 2001.

[13] Commission on Freedom of the Press. *A free and responsible press : a general report on mass communication: newspapers, radio, motion pictures, magazines, and books.* The University of Chicago Press, Chicago, IL, 1947.

[14] Federal Communications. *Report and Order and Notice of Proposed Rulemaking.* Federal Communications, Washington, DC, July 2003.

[15] Ann Coulter. *Slander: Liberal Lies about the American Right.* Three Rivers Press, September 2003.

[16] Stefano DellaVigna and Ethan Kaplan. The Fox News effect: Media bias and voting. *Quarterly Journal of Economics*, August 2007.

[17] David Dranove, Anne Gron, and Michael J. Mazzeo. Differentiation and competition in HMO markets. *Journal of Industrial Economics*, 51(4):433–454, December 2003.

[18] Jean-Pierre Dubé, Günter J. Hitsch, and Puneet Manchanda. An empirical model of advertising dynamics. *Quantitative Marketing and Economics*, 3(2):107–144, June 2005.

[19] Liran Einav. Seasonality in the US motion picture industry. *RAND Journal of Economics*, 2006. Forthcoming.

[20] Morris P. Fiorina, with Samuel J. Abrams, and Jeremy C. Pope. *Culture War? The Myth of a Polarized America*. Pearson Longman, New York, NY, 2005.

[21] C. Fox. A stop list for general text. *SIGIR FORUM*, 24(4), December 1990.

[22] Al Franken. *Lies and the Lying Liars Who Tell Them: A Fair and Balanced Look at the Right*. E P Dutton, August 2003.

[23] Wayne A. Fuller. *Measurement Error Models*. Wiley series in probability and mathematical statistics. Wiley, New York, 1987.

[24] Matthew Gentzkow. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*. Forthcoming.

[25] Matthew Gentzkow. Television and voter turnout. *Quarterly Journal of Economics*, 121(3), August 2006.

[26] Matthew Gentzkow and Jesse M. Shapiro. Media bias and reputation. *Journal of Political Economy*, 114(2):280–316, April 2006.

[27] Matthew A. Gentzkow, Edward L. Glaeser, and Claudia D. Goldin. The rise of the fourth estate: How newspapers became informative and why it mattered. In Edward L. Glaeser and Claudia Goldin, editors, *Corruption and Reform: Lessons from America's Economic History*, chapter 6, pages 187–230. University of Chicago Press, Chicago, IL, Spring 2006.

[28] Matthew A. Gentzkow and Jesse M. Shapiro. Media, education, and anti-Americanism in the Muslim world. *Journal of Economic Perspectives*, 18(3):117–133, Summer 2004.

[29] Lisa George. What's fit to print: The effect of ownership concentration on product variety in newspaper markets. *Michigan State University Mimeograph*, September 2001.

[30] Lisa George and Joel Waldfogel. Who affects whom in daily newspaper markets? *Journal of Political Economy*, 111(4):765–784, 2003.

[31] Alan Gerber, Dean Karlan, and Daniel Bergan. Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. *Yale University mimeograph*, February 2006.

[32] James G. Gimpel, Frances E. Lee, and Joshua Kaminski. The political geography of campaign contributions in American politics. *Journal of Politics*, 2006. Forthcoming.

[33] Edward L. Glaeser, Giacomo A. M. Ponzetto, and Jesse M. Shapiro. Strategic extremism: Why Republicans and Democrats divide on religious values. *Quarterly Journal of Economics*, 120(4):1283–1330, November 2005.

[34] Edward L. Glaeser and Bryce A. Ward. Myths and realities of American political geography. *Journal of Economic Perspectives*, 20(2):119–144, 2006.

[35] Theodore L. Glasser, David S. Allen, and S. Elizabeth Blanks. The influence of chain ownership on news play: A case study. *Journalism Quarterly*, 66:607–614, Fall 1989.

[36] Bernard Goldberg. *Bias : A CBS Insider Exposes How the Media Distort the News*. Perennial, January 2003.

[37] Michael J. Graetz and Ian Shapiro. *Death by a thousand cuts: The fight over taxing inherited wealth*. Princeton University Press, Princeton, NJ, February 2005.

[38] Tim Groseclose, Steven D. Levitt, and James M. Snyder, Jr. Comparing interest group scores across time and chambers: Adjusted ADA scores for the U.S. Congress. *American Political Science Review*, 93(1):33–50, March 1999.

[39] Tim Groseclose and Jeff Milyo. A measure of media bias. *Quarterly Journal of Economics*, 120(4):1191–1237, November 2005.

[40] S. I. Hayakawa. *Language in Thought and Action*. Harcourt Brace Jovanovich Inc., New York, 3rd (1972) edition, 1942.

[41] Harris Interactive. Seven in 10 U.S. adults say they watch broadcast news at least several times a week. *The Harris Poll*, 20, February 24 2006.

[42] Stephen Lacy and Todd F. Simon. Intercounty group ownership of daily newspapers and the decline of competition for readers. *Journalism and Mass Communication Quarterly*, 74(4):814–825, Winter 1997.

[43] Michael Laver, Kenneth Benoit, and John Garry. Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2):311–331, May 2003.

[44] John R. Lott, Jr. and Kevin A. Hassett. Is newspaper coverage of economic events politically biased? *American Enterprise Institute Mimeograph*, October 2004.

[45] Frank Luntz. *Learning from 2004...winning in 2006*. Luntz Research Companies, 2005.

[46] Hugh J. Martin. Some effects from horizontal integration of daily newspapers on markets, prices, and competition. *A paper presented to the Media Management and Economics Division of the Association for Education in Journalism and Mass Communication, Annual Convention, Kansas City, MO July 30-August 2, 2003*, July 2003.

[47] Michael J. Mazzeo. Product choice and oligopoly market structure. *RAND Journal of Economics*, 33(2):221–242, Spring 2002a.

[48] Michael J. Mazzeo. Competitive outcomes in product-differentiated oligopoly. *Review of Economics and Statistics*, 84(4):716–728, November 2002b.

[49] Sendhil Mullainathan and Andrei Shleifer. The market for news. *American Economic Review*, 95(4):1031–1053, September 2005.

[50] Caitlin Knowles Myers. Discrimination as a competitive device: The case of local television news. *IZA Discussion Paper No. 1802*, October 2005.

[51] Newspaper Association of America. The source: Newspapers by the numbers. *http://www.naa.org/thesource/6.asp*, 2006.

[52] David Pritchard. Viewpoint diversity in cross-owned newspapers and television stations: A study of news coverage of the 2000 presidential campaign. *FCC Media Ownership Working Group Working Paper*, September 2002.

[53] Riccardo Puglisi. Being the New York Times: the political behavior of a newspaper. *Massachusetts Institute of Technology Working Paper*, March 2006.

[54] Robert D. Putnam. *Bowling Alone: The Collapse and Revival of American Community*. Simon and Schuster, New York, 2000.

[55] Steven Ruggles, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. *Integrated Public Use Microdata Series: Version 3.0*. Minnesota Population Center, Minneapolis, MN, 2004. http://www.ipums.org.

[56] Katja Seim. An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics*, 2006. Forthcoming.

[57] Richard W. Stevenson. President makes it clear: Phrase is "war on terror". *New York Times*, page A.12, August 4 2005. Eric Schmitt contributed reporting to this article from Washington.

[58] James H. Stock and Motohiro Yogo. Testing for weak instruments in linear IV regression. *NBER Technical Working Paper 284*, 2002.

[59] David Strömberg. Radio's impact on public spending. *Quarterly Journal of Economics*, 119(1):189–221, 2004.

[60] Wing Suen. The self-perpetuation of biased beliefs. *Economic Journal*, 114:377–396, April 2004.

[61] Andrew Sweeting. Too much rock and roll? Station ownership, programming, and listenership in the music radio industry. *Northwestern University Mimeograph*, January 2006.

[62] US Supreme Court. *Associated Press v. United States*. US Supreme Court, 1945.

[63] David H. Weaver and G. Cleveland Wilhoit. *The American Journalist in the 1990s: U.S. News People at the End of an Era*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[64] Abigail K. Wozniak. Making their own luck: Educational differences in the migration responses of young workers to local labor market conditions. *University of Notre Dame Mimeograph*, January 2006.

**Table 1** *Politically loaded phrases from the 2005 Congressional Record*

Panel A: Phrases used more often by Democrats

*Two-word phrases*

| | | |
|---|---|---|
| private accounts | rosa parks | workers rights |
| trade agreement | president budget | poor people |
| american people | republican party | republican leader |
| tax breaks | change the rules | arctic refuge |
| trade deficit | minimum wage | cut funding |
| oil companies | budget deficit | american workers |
| credit card | republican senators | living in poverty |
| nuclear option | privatization plan | senate republicans |
| war in iraq | wildlife refuge | fuel efficiency |
| middle class | card companies | national wildlife |
| african american | security trust | president cheney |
| budget cuts | bill cuts | price gouging |
| nuclear weapons | medicaid cuts | iraq war |
| checks and balances | trade policy | million americans |
| civil rights | asian pacific | house republicans |
| veterans health | cia agent | assault weapons |
| cut medicaid | billions of dollars | senior citizens |
| foreign oil | abuse of power | cost of the war |
| president plan | manufacturing jobs | karl rove |
| gun violence | billion in tax | spending cuts |
| black caucus | lost their jobs | record profits |
| national debt | central american | bunker buster |
| public broadcasting | child labor | food stamps |
| child support | low income | bring our troops |
| student loans | cut programs | troops home |

*Three-word phrases*

| | | |
|---|---|---|
| veterans health care | corporation for public broadcasting | cut health care |
| congressional black caucus | additional tax cuts | civil rights movement |
| va health care | pay for tax cuts | cuts to child support |
| billion in tax cuts | tax cuts for people | drilling in the arctic national |
| credit card companies | oil and gas companies | victims of gun violence |
| security trust fund | prescription drug bill | solvency of social security |
| social security trust | caliber sniper rifles | voting rights act |
| privatize social security | increase in the minimum wage | war in iraq and afghanistan |
| american free trade | system of checks and balances | civil rights protections |
| central american free | middle class families | credit card debt |
| national wildlife refuge | cut student loans | little rock nine |
| dependence on foreign oil | american people deserve | social security plan |
| tax cuts for the wealthy | cut food stamps | arctic wildlife refuge |
| vice president cheney | health care education | education health care |
| arctic national wildlife | federal trade commission | social security the president |
| bring our troops home | congressional hispanic caucus | social security benefits |
| social security privatization | alternative minimum tax | explosive device detonated |
| billion trade deficit | asian and pacific islander | plan to privatize social |
| asian pacific american | global gag rule | ryan white care |
| president bush took office | cut social security | major oil companies |
| privatization of social security | billion in tax breaks | outing a cia agent |
| privatizing social security | below the poverty line | fuel economy standards |
| party line vote | middle class americans | improvised explosive device |
| child support enforcement | funding for veterans health | president social security |
| credit card industry | health care for veterans | international labor organization |

Source: Authors' calculations from the *Congressional Record*.

Notes: Table shows top words, ranked according to the $\chi^2$ statistic in a test of the independence between phrases and political party of the speaker. See section 3 for details.

Panel B: Phrases used more often by Republicans

*Two-word phrases*

| | | |
|---|---|---|
| stem cell | personal accounts | retirement accounts |
| natural gas | saddam hussein | government spending |
| death tax | pass the bill | national forest |
| illegal aliens | private property | minority leader |
| class action | border security | urge support |
| war on terror | president announces | cell lines |
| embryonic stem | human life | cord blood |
| tax relief | chief justice | action lawsuits |
| illegal immigration | human embryos | economic growth |
| date the time | increase taxes | food program |
| boy scouts | growth rate | time and i move |
| hate crimes | cell research | legal system |
| oil for food | property rights | nuclear power |
| global war | border patrol | democrat leader |
| medical liability | budget committee | growing economy |
| highway bill | consent decrees | raising taxes |
| adult stem | crimes law | witnesses may testify |
| democratic leader | post office | savings accounts |
| federal spending | european union | iraqi people |
| tax increase | president business | forest service |
| raise taxes | postal service | law we can change |
| illegal immigrants | terri schiavo | immigration reform |
| president i move | circuit court | indian affairs |
| third time | temporary worker | ten commandments |
| percent growth | war on terrorism | un reform |

*Three-word phrases*

| | | |
|---|---|---|
| embryonic stem cell | circuit court of appeals | tongass national forest |
| hate crimes legislation | death tax repeal | pluripotent stem cells |
| adult stem cells | housing and urban affairs | supreme court of texas |
| oil for food program | million jobs created | justice priscilla owen |
| personal retirement accounts | national flood insurance | justice janice rogers |
| energy and natural resources | oil for food scandal | american bar association |
| global war on terror | private property rights | growth and job creation |
| hate crimes law | temporary worker program | natural gas natural |
| change hearts and minds | class action reform | grand ole opry |
| global war on terrorism | chief justice rehnquist | reform social security |
| class action fairness | percent growth rate | judge john roberts |
| committee on foreign relations | united states postal service | supply of natural gas |
| deficit reduction bill | american farm bureau | gas natural gas |
| boy scouts of america | gross national product | chief of naval operations |
| repeal of the death tax | social security reform | underground storage tank |
| highway trust fund | export import bank | partial birth abortion |
| action fairness act | justice of the supreme court | judicial confirmation process |
| committee on commerce science | price of natural gas | personal savings accounts |
| cord blood stem | fifth circuit court | near earth objects |
| medical liability reform | social security system | national security issue |
| stem cell lines | committee on homeland security | law enforcement and intelligence |
| blood stem cells | united nations reform | justice william rehnquist |
| supreme court of the united | million illegal aliens | medical liability crisis |
| health savings accounts | california supreme court | judge alberto gonzales |
| banking housing and urban | term care insurance | economic growth and job |

Source: Authors' calculations from the *Congressional Record*.

Notes: Table shows top words, ranked according to the $\chi^2$ statistic in a test of the independence between phrases and political party of the speaker. See section 3 for details.

**Table 2** *Estimates of the demand for slant*

Dependent variable: Share of households in zipcode subscribing to newspaper

| Description | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Model | OLS | 2SLS | 2SLS/RC | 2SLS/RC | 2SLS/RC |
| (Zip share donating to Republicans) × Slant | 0.1733 (0.0740) | 0.6379 (0.1894) | 1.0897 (0.3165) | 0.8077 (0.2949) | 0.8505 (0.3119) |
| Zip share donating to Republicans | -0.0165 (0.0362) | -0.2281 (0.0879) | -0.4296 (0.1447) | -0.3251 (0.1380) | -0.3418 (0.1452) |
| (Zip share donating to Republicans)$^2$ | -0.0598 (0.0081) | -0.0615 (0.0079) | -0.0638 (0.0135) | -0.0353 (0.0129) | -0.0380 (0.0127) |
| Market-newspaper FE? | X | X | X | X | X |
| Zipcode demographics? | | | | X | X |
| Zipcode X market char.? | | | | | X |
| Estimate of $\alpha$ | 0.0954 | 0.3576 | 0.3942 | 0.4025 | 0.4019 |
| (Confidence interval) | (-1.17,0.30) | (0.21,0.40) | (0.30,0.43) | (0.25,0.44) | (0.25,0.44) |
| Estimate of $\beta$ | 0.6900 | 0.1929 | 0.1171 | 0.0874 | 0.0894 |
| (Confidence interval) | (0.32,3.06) | (0.11,0.47) | (0.06,0.29) | (0.02,0.34) | (0.02,0.34) |
| Estimate of $\gamma$ | 0.1256 | 1.6533 | 4.6547 | 4.6206 | 4.7553 |
| (Confidence interval) | (0.004,0.45) | (0.29,4.35) | (0.87,13.1) | (0.29,24.7) | (0.33,22.3) |
| Number of observations | 61845 | 61845 | 61845 | 61845 | 61845 |
| Number of newspapers | 290 | 290 | 290 | 290 | 290 |

Source: Authors' calculations based on Audit Bureau of Circulations (newspaper subscriptions), Federal Election Commission (campaign contributions), U.S. Presidential Atlas (county-level voting), U.S. Census (zipcode demographics), Editor and Publisher International Yearbook 2000-2005 (newspaper location).
Notes: Table shows estimates of models of the form of equation (9). Standard errors (in parentheses) are clustered by newspaper. Zipcode demographics are log of total population, log of income per capita, percent of population urban, percent white, percent black, population per square mile, share of houses owner-occupied, and the share of population 25 and over whose highest level of schooling is college, all as of 2000. "Zipcode X market characteristics" refers to a vector of these characteristics interacted with their analogue at the level of the newspaper's market.

**Table 3** *Ownership and newspaper slant*

Dependent variable: Slant index ($\hat{y}_n$)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Average slant of other newspapers in ownership group | 0.6040 (0.1159) | 0.5453 (0.1375) | 0.4217 (0.1843) | 0.2438 (0.2139) |
| Control for profit-maximizing slant? | | X | X | X |
| Census division fixed effects? | | | X | |
| State fixed effects? | | | | X |
| Number of observations | 338 | 338 | 338 | 338 |
| Number of ownership groups | 36 | 36 | 36 | 36 |
| $R^2$ | 0.0877 | 0.0713 | 0.0393 | 0.0130 |

Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), U.S. Presidential Atlas (county-level voting), Editor and Publisher International Yearbook 2000-2005 (newspaper location and ownership).

Notes: Standard errors (in parentheses) clustered by ownership group. See section 3 for derivation of slant index and section 4.4 for details on calculation of profit-maximizing slant. In specifications (2) through (4), slant index is regressed on controls, and then residuals are averaged to form adjusted average slant of other newspapers in ownership group.

**Table 4** *Decomposing the variation in newspaper slant*

Dependent variable: Slant index ($\hat{y}_n$)

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Profit-maximizing slant in newspaper's market | — | 2.0340 (0.2413) | 1.9136 (0.1930) | 2.1078 (0.2029) | 2.2246 (0.2039) |
| Ownership group fixed effects? | | X | | | |
| Census division fixed effects? | | | | X | |
| State fixed effects? | | | | | X |
| Standard deviation of ownership effect | 0.0144 (0.0034) | | 0.0121 (0.0039) | 0.0046 (0.0032) | 0.0000 (0.0051) |
| Ownership share of residual variation | 0.1324 (0.0633) | | 0.0943 (0.0529) | 0.0208 (0.0206) | 0.0000 (0.0599) |
| Consumer share of residual variation | | 0.1910 (0.0453) | 0.2005 (0.0404) | 0.2071 (0.0399) | 0.2238 (0.0410) |
| Number of observations | 413 | 413 | 413 | 413 | 413 |
| Number of multi-paper groups | 36 | 36 | 36 | 36 | 36 |

Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), U.S. Presidential Atlas (county-level voting), Editor and Publisher International Yearbook 2000-2005 (newspaper location and ownership).

Notes: See section 3 for derivation of slant index. Newspaper market is defined as the newspaper's primary metropolitan statistical area if available, and the newspaper's county if not. Models estimated via maximum likelihood. Standard errors on the standard deviation of the ownership effect and the ownership share of the variation are obtained through a parametric bootstrap. Ownership and consumer share of residual variation are the share of variation in slant explained by ownership group random effects and profit-maximizing slant respectively; in columns (2), (4) and (5) the share(s) are computed after partialling for group, division, and state fixed effects respectively.

**Table 5** *Robustness of the relationship between slant and consumer characteristics*

Dependent variable: Slant index ($\hat{y}_n$)

| | (1) 2SLS | (2) 2SLS | (3) OLS | (4) OLS |
|---|---|---|---|---|
| Instrument(s) | % church | log population, % black, % college | | |
| Profit-maximizing slant in newspaper's market | 1.8565 (0.7609) | 3.6437 (0.3642) | 1.0654 (0.1955) | 1.2073 (0.1942) |
| Log(market population) (2000) | | | -0.0057 (0.0012) | -0.0014 (0.0015) |
| Share black in market (2000) | | | -0.1471 (0.0149) | -0.1408 (0.0147) |
| Share college-educated in market (2000) | | | -0.0530 (0.0247) | -0.0304 (0.0247) |
| Log(number of newspaper employees) | | | | -0.0023 (0.0022) |
| Log(number of pages) | | | | -0.0133 (0.0052) |
| Number of Pulitzers, 1970-2006 | | | | -0.0004 (0.0005) |
| Number of observations | 406 | 413 | 413 | 413 |
| $R^2$ | — | — | 0.4231 | 0.4560 |

Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), U.S. Presidential Atlas (county-level voting), Editor and Publisher International Yearbook 2000-2005 (newspaper location, ownership, and characteristics), DDB Needham LifeStyle survey 1972-1998 (church attendance), U.S. Census 2000 (demographics), <www.pulitzer.org> (number of Pulitzer prizes).

Notes: Standard errors in parentheses. See section 3 for derivation of slant index and section 4.4 for details on calculation of profit-maximizing slant. Specification (1) uses the share attending church monthly from 1972-1998 in the newspaper's primary market as an instrument for ideal slant. Specification (2) uses log population, share black, and share with a college degree in the newspaper's primary market as instruments for slant. Number of employees and number of pages are reported in the 2001 *Editor and Publisher International Yearbook*. In column (4), dummies are included to control for missing values of number of employees and number of pages.

**Table 6** *The response of slant to financial incentives*

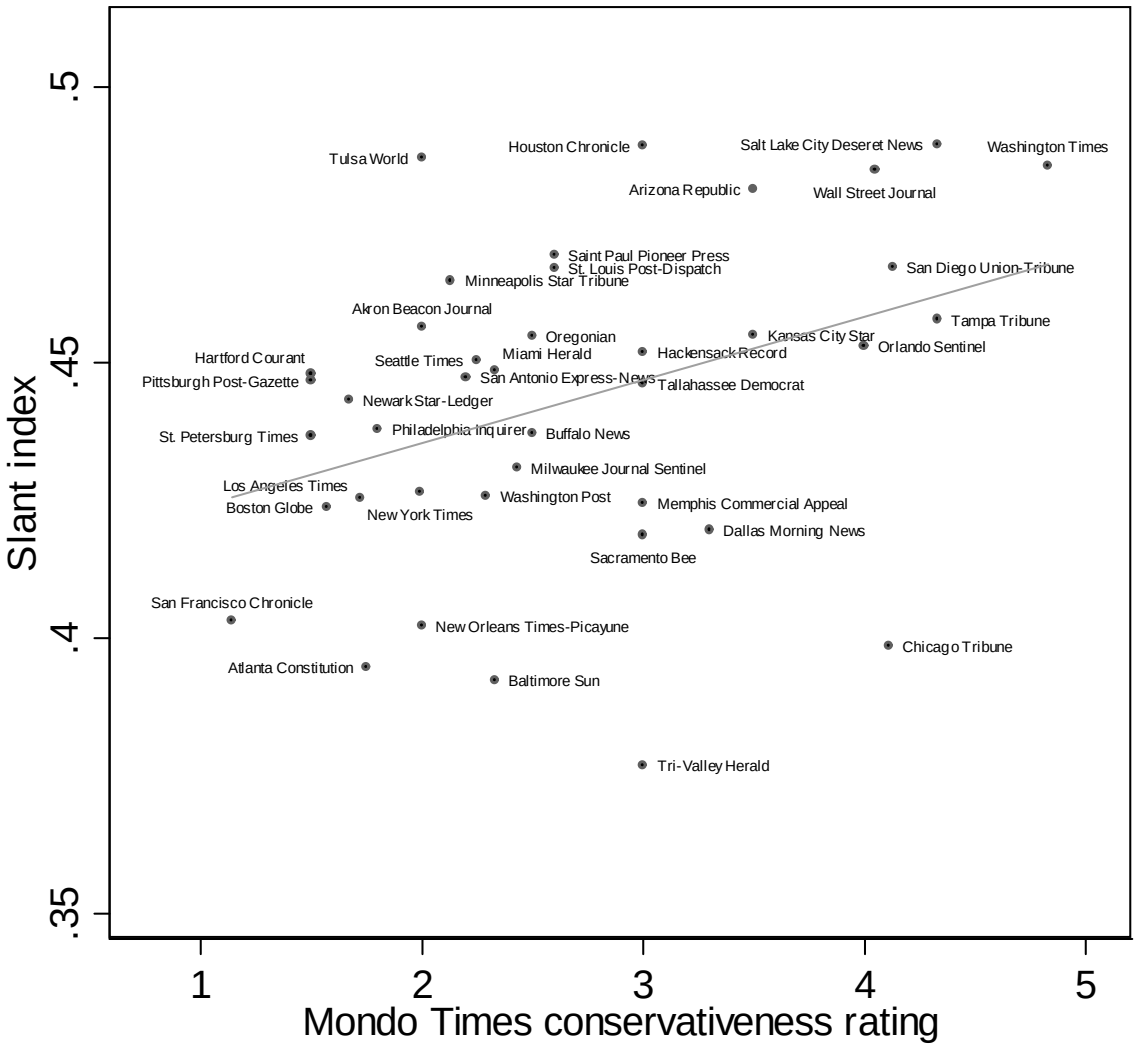
Dependent variable: Slant index ($\hat{y}_n$)

| Financial variable: | Advertising rate per reader | | Ownership structure | |
|---|---|---|---|---|
| Sample | Below-median | Above-median | Private | Public |
| Profit-maximizing slant in newspaper's market | 1.6311 (0.2742) | 1.7487 (0.2708) | 2.3161 (0.2628) | 1.2858 (0.3775) |
| *Difference in coefficients* | *0.1175 (0.3791)* | | *-1.0302 (0.4605)* | |
| Standard deviation of ownership effect | 0.0095 (0.0045) | 0.0152 (0.0046) | 0.0119 (0.0053) | 0.0174 (0.0065) |
| *Difference in standard deviations* | *0.0056 (0.0056)* | | *0.0055 (0.0082)* | |
| Number of observations | 395 | 395 | 357 | 357 |

Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), U.S. Presidential Atlas (county-level voting), Editor and Publisher International Yearbook 2000-2005 (newspaper location, ownership, and advertising rates), various sources (ownership structure).

Notes: Standard errors in parentheses. See section 3 for derivation of slant index and section 4.4 for details on calculation of profit-maximizing slant. Models estimated via maximum likelihood, with the effect of the owner-level random component permitted to vary with the financial variable listed. A public firm is defined as a firm that is publicly traded, in which no single shareholder or family has a majority interest.

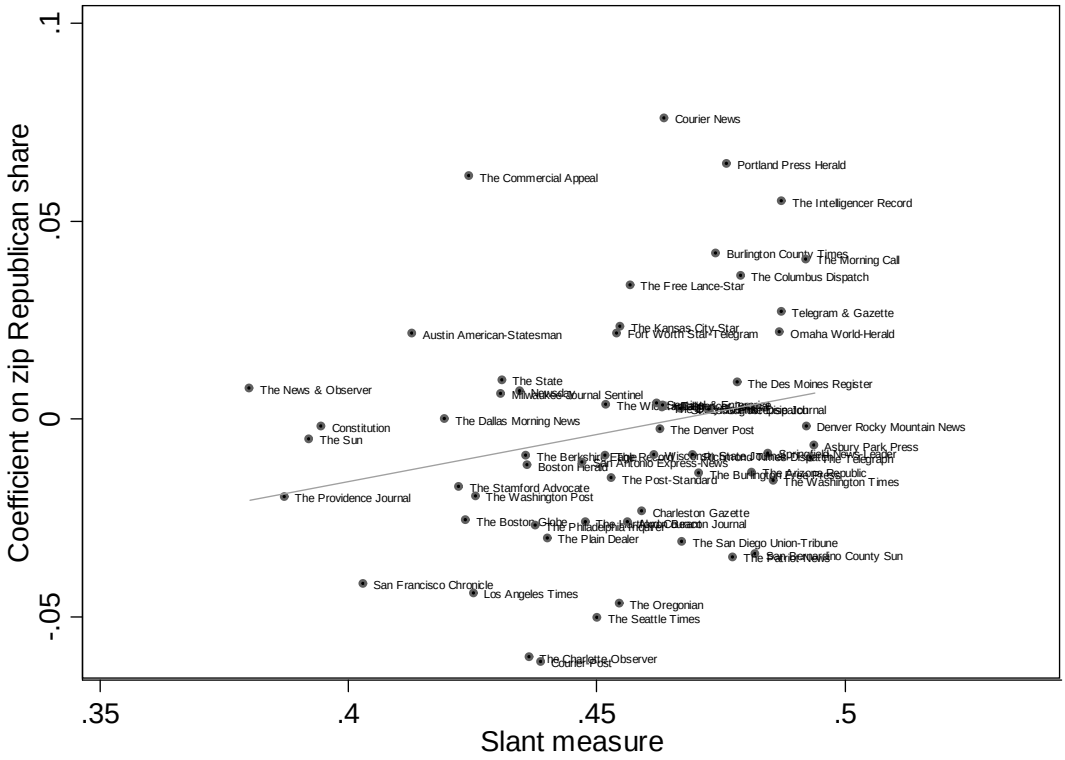
**Figure 1** *Language-based and reader-submitted ratings of slant*




Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), Mondo Times at <http://www.mondotimes.com> (bias ratings).

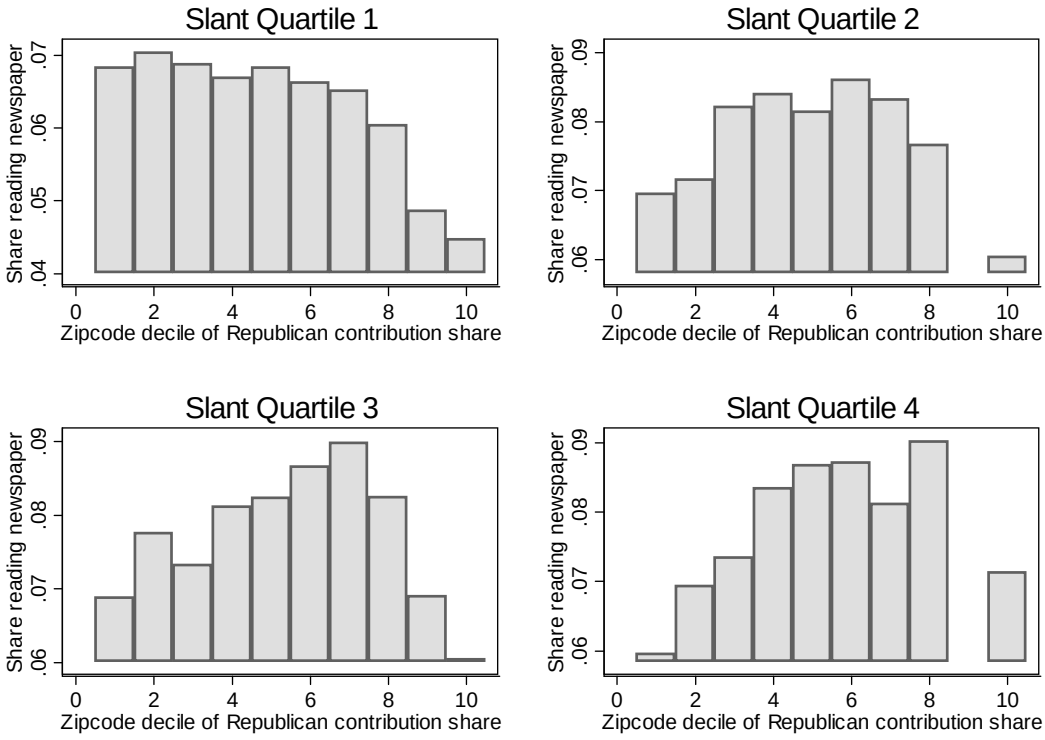
Notes: Figure shows slant index (y-axis) against average Mondo Times user rating of newspaper conservativeness (x-axis), which ranges from 1 (liberal) to 5 (conservative). See section 3 for derivation of slant index. Figure includes all papers rated by at least two users on Mondo Times, with at least 25,000 mentions of our 1,000 phrases in 2005.

**Figure 2** *Newspaper slant and consumer demand*



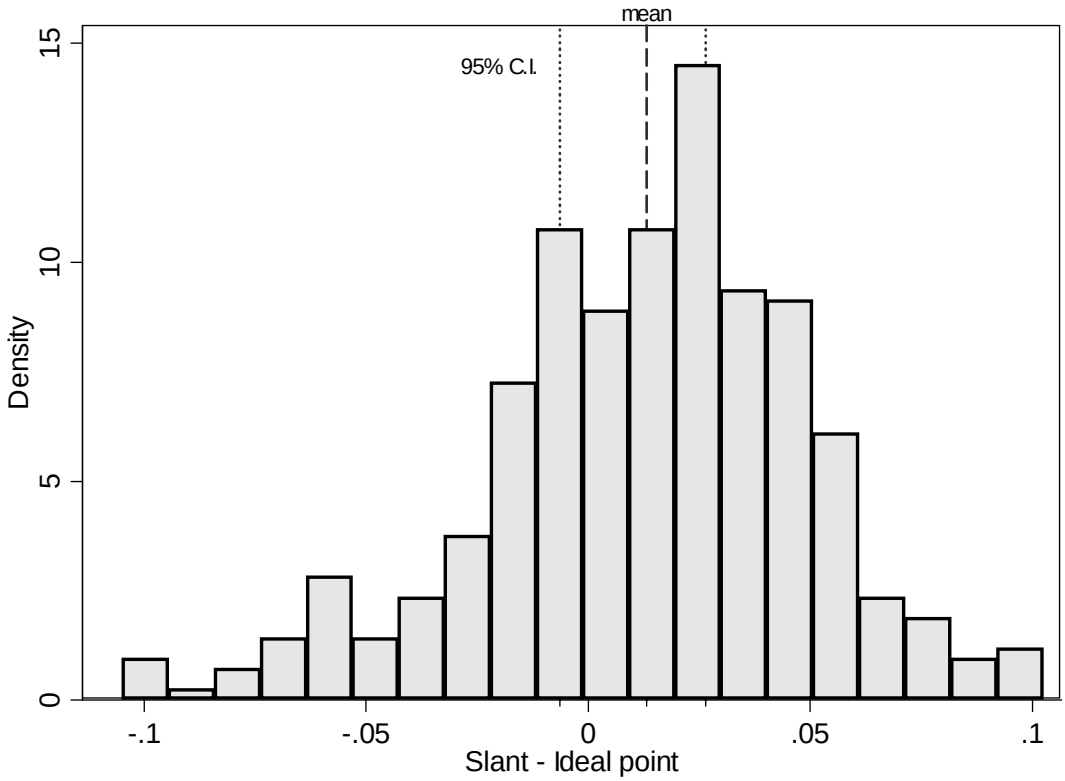

Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), Audit Bureau of Circulations (newspaper subscriptions), Federal Election Commission (campaign contributions)
Notes: Y-axis shows the estimated effect of the share contributing to Republican candidates on the share of households in the zipcode reading each newspaper, from a model in which readership shares are regressed, separately by newspaper, on contribution shares and market fixed effects. X-axis shows slant measure. Figure excludes data for newspapers circulating in fewer than 300 zipcodes.

**Figure 3** *Newspaper slant and variation in consumer demand*



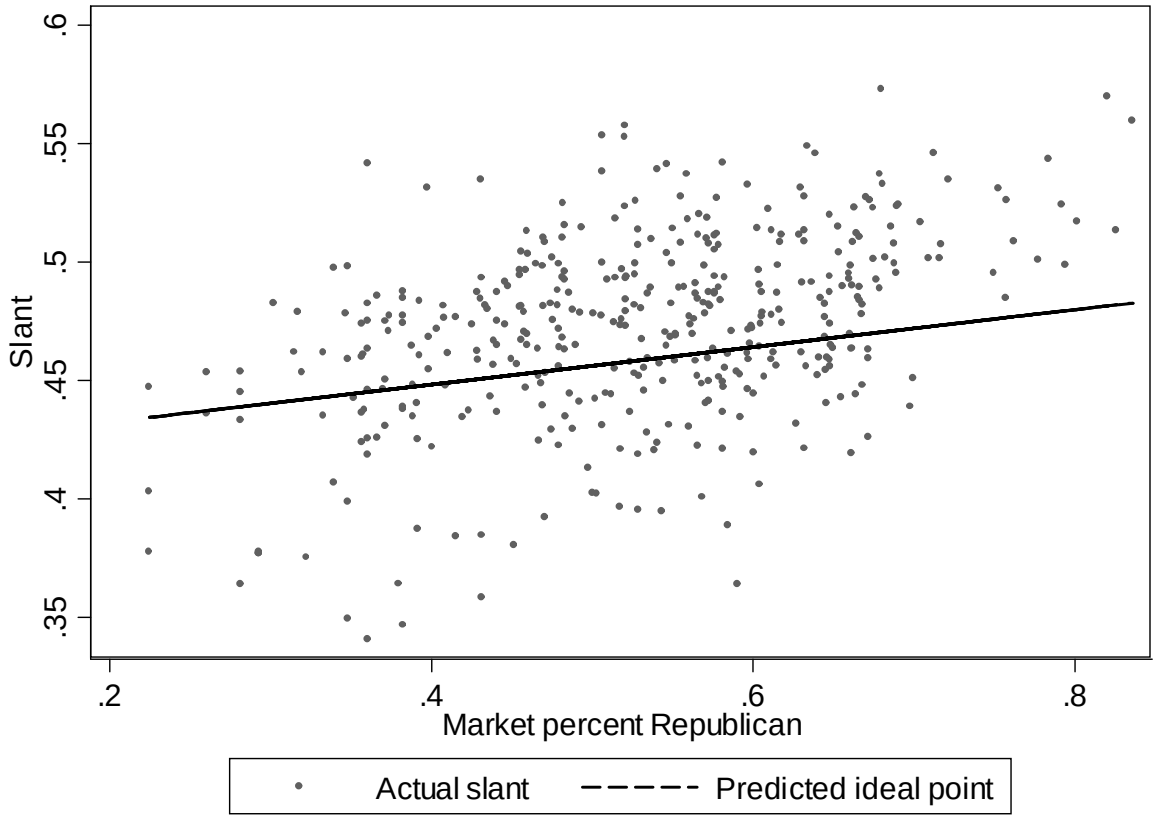

Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), Audit Bureau of Circulations (newspaper subscriptions), Federal Election Commission (campaign contributions).
Notes: Figure shows coefficients on decile dummies in regressions of the share of households in a zipcode reading a newspaper on dummies for decile of share donating to Republicans in the 2000-2004 election cycle, with market-newspaper fixed effects, and weighted by zipcode population. Equation is estimated separately for newspapers in each quartile of the distribution of measured slant.

**Figure 4** *Differences between slant and predicted ideal point*




Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), U.S. Presidential Atlas, FEC contribution data, and Audit Bureau of Circulations (ideal points).
Notes: Figure shows the distribution of the difference between newspapers' actual slant and our estimate of their profit-maximizing level of slant ($\hat{y}_n - ideal_n$). See section 3 for derivation of slant index, and section 4.4 for details on the computation of profit-maximizing level of slant. The dashed line indicates the mean of the distribution and the dotted lines indicate the 95 percent confidence interval for the value of the mean (incorporating both sampling variation in slant and uncertainty in the demand estimates that are inputs to computing $ideal_n$).

**Figure 5** *Slant and consumer preferences*




Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), U.S. Presidential Atlas (county-level voting), Editor and Publisher International Yearbook 2000-2005 (newspaper location).

Notes: Figure shows newspaper slant index and profit-maximizing level of slant (y-axis) against Bush's share of the two-party vote in 2004 in the newspaper's market (x-axis). See section 3 for derivation of slant index, and section 4.4 for details on the computation of profit-maximizing level of slant. Newspaper market is defined as the newspaper's primary metropolitan statistical area if available, and the newspaper's county if not.

**Figure 6** *Newspaper slant and ownership*

Figure A: Relationship between newspaper slant and average slant of co-owned papers

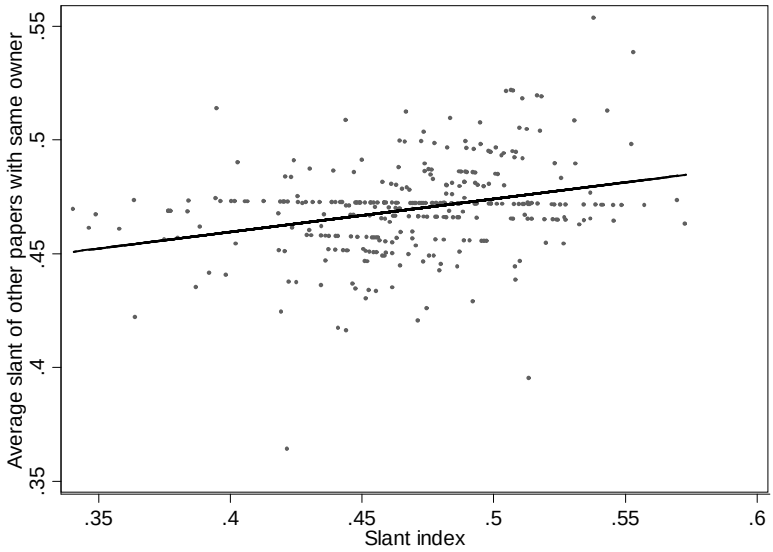
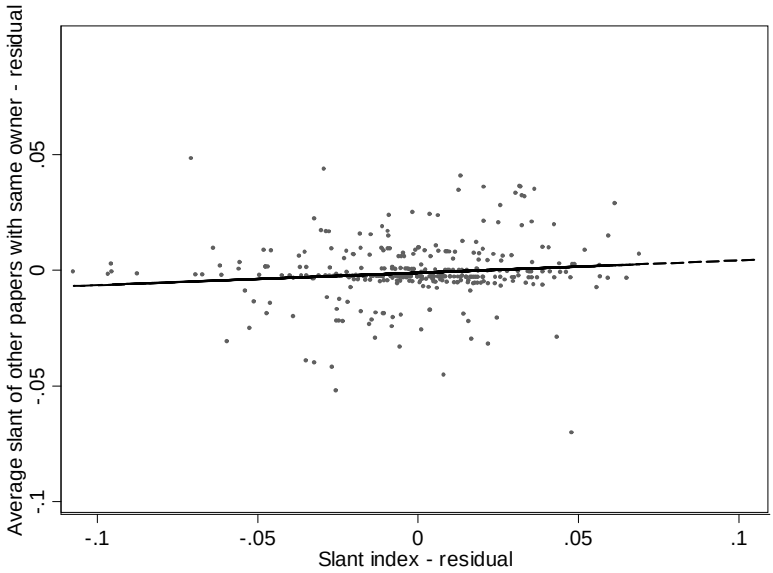
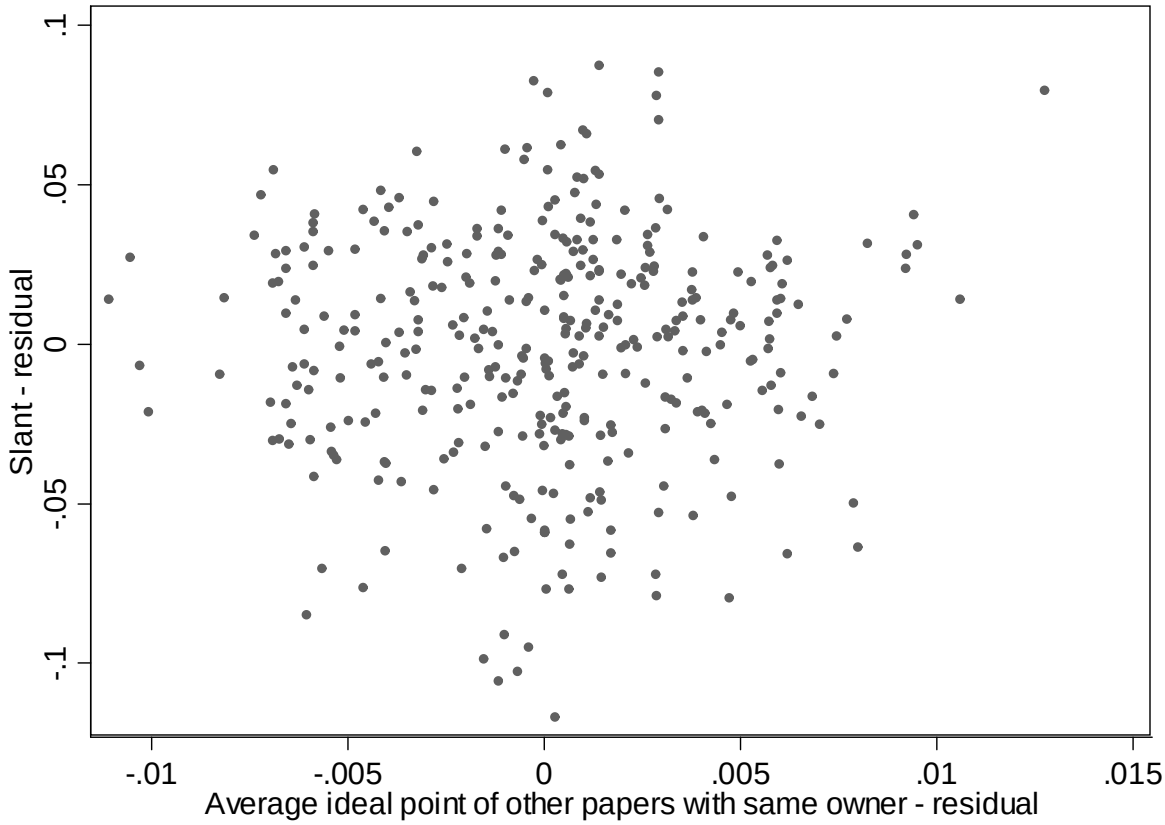



Figure B: Newspaper slant and slant of co-owned papers, controlling for consumer preferences and state




Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), Editor and Publisher International Yearbook 2000-2005 (newspaper location and ownership).
Notes: See section 3 for derivation of slant index and section 4.4 for details on calculation of profit-maximizing slant. Figure A shows average slant of co-owned newspapers graphed against a newspaper's own slant. Figure B parallels figure A, but measures slant using residuals from a regression of slant on profit-maximizing slant and dummies for the state in which the newspaper is located.

**Figure 7** *Testing for fixed costs in the production of news content*




Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), Editor and Publisher International Yearbook 2000-2005 (newspaper location and ownership).
Notes: Both variables partialled with respect to the profit-maximizing level of slant in the newspaper's market. See section 3 for derivation of slant index, and section 4.4 for details on the computation of profit-maximizing level of slant. Newspaper market is defined as the newspaper's primary metropolitan statistical area if available, and the newspaper's county if not.

**Figure 8** *Newspaper slant and political contributions*

Figure A: Newspaper slant and donations of top-ranking corporate executives and officers

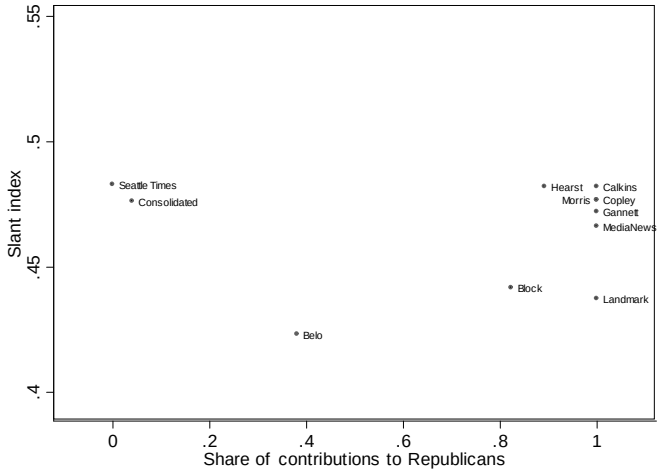
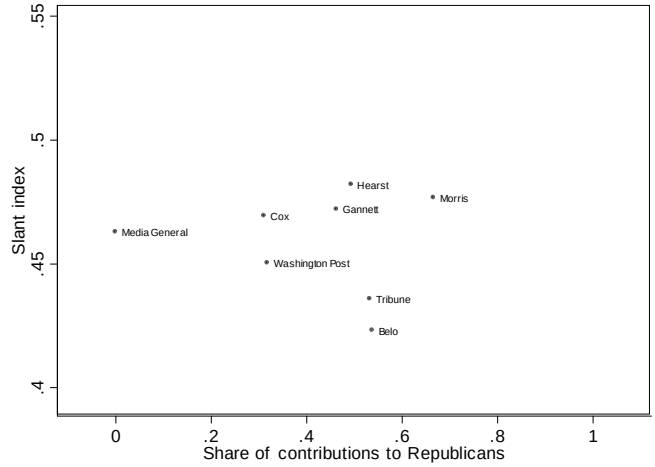



Figure B: Newspaper slant and corporate donations




Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), Editor and Publisher International Yearbook 2000-2005 (newspaper ownership), Federal Election Commission (donations of executives), Center for Public Integrity (corporate donations).
Notes: Figure A shows average slant of newspapers owned by a firm graphed against the share of total contribution dollars going to Republicans from the CEO, President, Managing Director, or Chairman of the Board, as collected from the FEC's disclosure database. Figure B shows average slant graphed against the share of corporate contribution dollars going to Republicans, as measured by the Center for Public Integrity.

**Appendix Table 1** *Additional robustness checks*

| Specification | | Profit-maximizing slant in newspaper's market | Ownership share of residual variation | Consumer share of residual variation |
|---|---|---|---|---|
| (1) | Baseline | 2.2246 | 0.0000 | 0.2238 |
| | | (0.2039) | (0.0599) | (0.0410) |
| (2) | Logit demand model | 2.2325 | 0.0000 | 0.2238 |
| | | (0.2046) | (0.0599) | (0.0410) |
| (3) | Logit demand model with cross-paper substitution | 2.1679 | 0.0000 | 0.2238 |
| | | (0.1987) | (0.0599) | (0.0410) |
| (4) | Exclude newspapers in multi-paper cities | 2.0099 | 0.0000 | 0.1753 |
| | | (0.2150) | (0.0336) | (0.0375) |
| (5) | Controlling for predicted sophistication | 2.2270 | 0.0000 | 0.2243 |
| | | (0.2056) | (0.0598) | (0.0414) |
| (6) | Tightening cutoffs on phrase counts by 5% | 3.5729 | 0.0000 | 0.1942 |
| | | (0.3581) | (0.0553) | (0.0389) |
| (7) | Measuring ideology with adjusted ADA score | 1.8389 | 0.0000 | 0.2009 |
| | | (0.1805) | (0.0246) | (0.0394) |

Source: Authors' calculations based on ProQuest database and *Congressional Record* (slant index), U.S. Presidential Atlas (county-level voting), Editor and Publisher International Yearbook 2000-2005 (newspaper location and ownership).

Notes: Standard errors in parentheses. See section 3 for derivation of slant index, and section 4.4 for details on the computation of profit-maximizing level of slant. Newspaper market is defined as the newspaper's primary metropolitan statistical area if available, and the newspaper's county if not. Models include state fixed effects and owner random effects, and are estimated via maximum likelihood. Standard errors on the ownership share of the variation are obtained through a parametric bootstrap. Ownership and consumer share of residual variation are the share of variation in slant explained by ownership group random effects and profit-maximizing slant respectively; these shares are computed after partialling for state fixed effects. See appendix A for details.