

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 Attila Ambrus, David Autor, Gary Becker, Gary Chamberlain, Raj Chetty, Tim Conley, Liran Einav, Edward Glaeser, Tim Groseclose, Christian Hansen, Justine Hastings, Chris Hayes, Daniel Hojman, Matt Kahn, Larry Katz, John List, Kevin M. Murphy, Ben Olken, Ariel Pakes, Andrea Prat, Riccardo Puglisi, Sam Schulhofer-Wohl, Andrei Shleifer, Monica Singhal, Jim Snyder, Wing Suen, Catherine Thomas, Abe Wickelgren, and numerous seminar and conference participants 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, Lisa Furchtgott, Ingrid Gonçalves, Hayden Haralson Hudson, and Hannah Melnicoe for excellent research assistance. This research was supported by National Science Foundation Grant SES-0617658 and the Centel Foundation/Robert P. Reuss Faculty Research Fund, the George J. Stigler Center for the Study of the Economy and the State, and the Initiative on Global Markets, all at the University of Chicago Graduate School of Business. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2006 by Matthew Gentzkow and Jesse M. Shapiro. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

What Drives Media Slant? Evidence from U.S. Daily Newspapers  
Matthew Gentzkow and Jesse M. Shapiro  
NBER Working Paper No. 12707  
November 2006, Revised August 2007  
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 that of 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. We also present evidence on the role of pressure from incumbent politicians, tastes of reporters, and newspaper competition in determining slant.

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).<sup>1</sup>

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).<sup>2</sup> 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

---

<sup>1</sup>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.’”

<sup>2</sup>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.”

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, 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 to test 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.<sup>3</sup> We focus on newspapers' news (rather than opinion) content, because of its centrality to public policy debates and its importance as a source of information to consumers.<sup>4</sup> To measure news slant, 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. Note that our measure has no inherent normative content—in our framework, one paper's slant can fall to the left or right of another's, but the notion that a paper has “less slant” or is “less biased” has no meaning.

Two key pieces of evidence suggest that our methodology produces a meaningful measure of slant. First, many of the phrases that our automated procedure identi-

---

<sup>3</sup>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.

<sup>4</sup>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).

fies 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).

We use our measure to estimate a Hotelling model of newspaper demand, in which a consumer’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 a higher proportion of Republicans, even within a narrowly defined geographic market. This fact is robust to correcting for measurement error and endogeneity in newspaper slant using an identification strategy in the spirit of George and Waldfogel (2003). 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 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 neither to the right nor to the left of the profit-maximizing level on average. Although we replicate Groseclose and Milyo’s (2005) finding that the average newspaper’s language is similar to that of a left-of-center member of Congress, we estimate that the profit-maximizing points are also left-of-center on average. These findings are relevant to theories in which supply-side forces cause distortions in slant at the aggregate level. For example, if either the party identity of national incumbent politicians (Besley and Prat 2006) or the distribution of political views among journalists in the country as a whole (Baron 2006) were important drivers of slant, we would have expected to see deviation from profit maximization on average.

Our second finding is that the variation in slant across newspapers is strongly related to the political makeup of their potential readers, and thus to our estimated profit-maximizing points. The relationship between slant and consumer characteristics remains when we compare different newspapers with the same owner, or different newspapers located in the same state, and survives controlling for measures of newspaper quality. Changes over time in consumer politics are a highly significant predictor of changes in newspaper slant. Overall, variation in consumer political attitudes explains roughly 20 percent of the variation in measured slant in our sample.

An obvious concern in interpreting the relationship between slant and consumer attitudes is that it may reflect causation running from slant to consumer beliefs rather than the reverse. To address this, we show that the relationship survives when we instrument for consumer political attitudes using characteristics such as religiosity and race that are strong predictors of political preferences but are unlikely to be affected by newspaper content. These results do not mean that newspapers do not affect beliefs—indeed, our study is motivated in part by evidence that they do. Rather, our findings suggest that the effect of slant on ideology accounts for only a small part of the cross-sectional variation in ideology that identifies our model.

Our third finding is that ownership plays little or no role in determining slant. After controlling for geographic clustering of newspaper ownership groups, the slant of co-owned papers is only weakly (and statistically insignificantly) related to 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. We also find no evidence that owner effects arise when political incentives are strong (i.e., during close elections), and no evidence of systematic changes in slant around mergers. Finally, direct proxies for owner ideology, such as patterns of corporate or executive donations to political parties, are unrelated to slant.

In the final section of the paper, we consider evidence on three additional forces that may affect slant—pressure from incumbent politicians (Besley and Prat 2006), the tastes of reporters and editors (Baron 2006), and competition (Mullainathan and Shleifer 2005). As mentioned above, our finding that newspapers are neither to the right nor to the left of our estimated ideal points on average suggest that the importance of the first two of these forces is limited at the aggregate level. We go

on to ask whether political pressure or reporters tastes can explain the *variation* in slant across papers, and we present several pieces of evidence suggesting that they are unlikely to play a large role. We find evidence consistent with Mullainathan and Shleifer’s (2005) prediction that more competitive media markets will have greater cross-outlet variation in slant, but our test of this hypothesis has limited power.

We wish to stress two important caveats. First, our results are specific to the context we are studying—a developed economy with a relatively competitive press and strong protections for press freedom. The picture will almost certainly look different in other settings, and understanding how the drivers of slant change with the legal and institutional environment is an important topic of ongoing research. Second, our measure of slant is a broad aggregate that includes coverage of many different topics over a reasonably long window of time. Nothing we find precludes the possibility that owners, politicians, or reporters may exert significant influence on coverage of specific stories where their incentives are especially high.

This paper presents the first large-scale empirical evidence on the determinants of political slant in the news.<sup>5</sup> Hamilton (2004) presents an important overview of many of the issues we explore. 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 role of political pressure and reporters’ tastes 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; Larcinese, Puglisi and Snyder 2006; Gentzkow, Glaeser, and Goldin 2006).<sup>6</sup> Groseclose and Milyo (2005) use Congressional citations to estimate

---

<sup>5</sup>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 (and vice versa).

<sup>6</sup>Our approach borrows tools from the computer science literature on “text categorization” (see

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 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. 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 tests several prominent theories about the determinants of media slant. Section 7 concludes.

## 2 Data

### 2.1 *Congressional Record* and Congressperson Data

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

---

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 our knowledge) to the political slant of the news media.



*sional Record*,<sup>7</sup> parsed using an automated script that identifies the speaker of each passage.<sup>8</sup>

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).<sup>9</sup> A final script produces counts by speaker and party of two- and three-word phrases in the *Congressional Record*.<sup>10</sup>

For each congressperson,<sup>11</sup> 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).<sup>12</sup> We also obtain data on the state, Census division, and demographic characteristics of each congressperson’s constituency.<sup>13</sup>

---

<sup>7</sup>Our database of Congressional Record text is incomplete, mostly due to errors in the website that archives the Congressional Record. These errors affect a relatively small share of documents in the Congressional Record (roughly 15 percent).

<sup>8</sup>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.

<sup>9</sup>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.

<sup>10</sup>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.

<sup>11</sup>We use the word “congressperson” as a generic term to refer to members of both the House of Representatives and the Senate.

<sup>12</sup>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.

<sup>13</sup>Data on presidential vote shares and demographic characteristics of congressional districts are from <[www.polidata.org](http://www.polidata.org)>.

## 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 and the ProQuest Newsstand database.<sup>14</sup> For each database, we use an automated script to calculate the number of articles containing each phrase in each newspaper during calendar year 2005. Whenever possible, we exclude opinion content. Also, because some newspapers do not archive reprinted wire stories with ProQuest, we exclude articles from the Associated Press, focusing instead on content originating with the newspaper. Appendix A provides additional details on the mechanics of these searches, as well as an audit study of the extent to which we successfully exclude opinion and wire content.

We include in our sample only English-language daily newspapers. Data are available for 394 such newspapers from NewsLibrary and 164 from ProQuest, with an overlap of 125 newspapers. This leaves us with a total sample of 433 newspapers in 2005.<sup>15</sup> Among the newspapers that overlap between the two databases, the correlation between the counts for the sample of phrases described below is .85, indicating high cross-database reliability.<sup>16</sup> 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 2001 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

---

<sup>14</sup>See <<http://www.newslibrary.com>> for the NewsLibrary database and <<http://proquest.com>> for the ProQuest database.

<sup>15</sup>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. A large share of hits for this paper are for a single phrase, “African American,” which is strongly predictive of liberal ideology in Congress.

<sup>16</sup>Possible reasons that the databases do not agree perfectly include: (i) a lag between the publication of an article and its posting; (ii) differences in the set of articles newspapers choose to post to each database; and (iii) differences in how the two databases permit us to identify editorials and opinion pieces.

be the county in which it is located.

For each newspaper, we obtain a wide range of demographic characteristics of the paper’s market from the 2000 U.S. Census. 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 the DDB Needham Life Style Survey (Putnam 2000), available on [www.bowlingalone.com](http://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](http://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.

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 firm’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 U.S. daily newspapers. For newspapers owned by a firm with no other daily newspaper holdings, we conducted an analogous search, but collected data on executives of the newspaper itself.

### **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 297 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.<sup>17</sup>

---

<sup>17</sup>Because our analysis will use only variation across zipcodes within a market, such cases provide no additional variation with which to identify our models.

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 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 (denominated in number of donations, not dollars) 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. To test its validity, we take advantage of data on the number of registered Democrats and Republicans by zipcode in California as of March 2006.<sup>18</sup> A zipcode-level regression of our donation measure on the share of Republican registrants, weighted by the total number of donors in the zipcode, has an  $R^2$  of 0.65. The donation and registration variables are also highly correlated across U.S. counties.

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

---

<sup>18</sup>We are grateful to Marc Meredith for providing these data.

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.<sup>19</sup> In 2005, the *Post* used the phrase “estate tax” 13.7 times as often as it used the phrase “death tax,” while the *Times* used “estate tax” 1.3 times as often. As we show below, this case is not unusual: there is a significant 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

---

<sup>19</sup>The website <[www.mondotimes.com](http://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*.

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 Democrats and Republicans:

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

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.<sup>20</sup> 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

---

<sup>20</sup>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.

than 400,000 documents.<sup>21</sup> 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.<sup>22</sup> Panel A shows phrases used more often by congressional Democrats. Panel B 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

---

<sup>21</sup>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 B for details.) Our findings therefore do not seem particularly sensitive to the choice of headline count cutoffs for these phrases.

<sup>22</sup>Some of the two-word phrases on our list are proper subsets of three-word phrases that also appear on the list. These phrases are not necessarily redundant statistically, because our preprocessing step (removal of stopwords and stemming) means that the two-word phrases may arise in somewhat different contexts than the three-word phrases. Nevertheless, we have verified that excluding these phrases produces a slant measure highly correlated with our own.

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.”

### 3.2 Mapping Phrases to Ideology

The next step is to map the use of these 1,000 phrases into a continuous measure of ideology. For computational simplicity, the  $\chi^2$  statistic we used to select the phrases was based only on whether a speaker was a Republican or a Democrat. To increase precision, we now wish to use a measure that will distinguish relatively liberal or conservative members of each party. The measure we will use is the share  $y_c$  of voters in the congressperson  $c$ ’s constituency voting for the Republican presidential candidate in 2004.<sup>23</sup> For each congressperson  $c$ , and phrase  $p$ , we define  $s_{pc}$  to be 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-meanned” 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-meanned) frequencies  $\tilde{s}_{pn}$  for individual

---

<sup>23</sup>When we instead measure ideology using adjusted ADA scores based on congressional roll call votes, the resulting slant estimate is highly correlated with our baseline measure, and produces very similar results in the analyses of newspaper slant we report below. See appendix B for details.



$n$ . That is, we will choose  $\bar{y}_n$  to solve

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

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

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

This estimator is interpretable as a sum of the phrase frequencies  $\tilde{s}_{pn}$ , scaled by the relationship between the ideology of a congressperson and the frequency with which that congressperson uses the phrase.<sup>24</sup> 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  $\bar{y}_n$ . 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 congressional roll-call voting behavior (as measured by adjusted ADA scores). These findings lend support to the expectation that our estimator will uncover genuine variation in ideological slant among newspapers, although they also indicate that our measure contains a nontrivial amount of noise.

---

<sup>24</sup>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  $y_c$ .

### 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 opinion pieces. Our search also excludes Associated Press stories.

Despite these exclusions, our 1,000 phrases are used an average of over 13,000 times in the content of papers in our sample in 2005. Even among newspapers in the bottom quartile of daily circulation in our sample, these phrases are used an average of over 4,000 times. The contexts in which our phrases appear include local analogues of national issues, local impact of federal legislation, and the actions of legislators from local districts.<sup>25</sup>

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>).<sup>26</sup> 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 103 papers in our sample rated by more than one individual on the Mondo Times website, there is a correlation of 0.24 with our slant index ( $p$ -value = 0.015), and a rank correlation of 0.26 ( $p$ -value = 0.007). 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 opinion as well as news content of the papers, whereas our slant measure focuses on news content.

---

<sup>25</sup>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.

<sup>26</sup>We wish to thank Eric Kallgren of Mondo Code for graciously providing these data.

We can also compare our measure to Groseclose and Milyo’s (2005) bias measures for the six newspapers that were part of their analysis.<sup>27</sup> Like Groseclose and Milyo (2005), we find that the *New York Times*, *Los Angeles Times*, and *Washington Post* are similar to one another and are well to the left of the *Washington Times*, which both measures identify as the most right-leaning newspaper of this group. Our measures disagree on two points: (i) our measure places *USA Today* just to the left of the *New York Times*, *Los Angeles Times*, and *Washington Post*, whereas their measure identifies it as more centrist than those papers; and (ii) we identify the *Wall Street Journal* as the second most right-leaning of these newspapers, whereas Groseclose and Milyo (2005) estimate that it is the most liberal.

## 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 model as a linear function of the zipcode’s Republicanism

---

<sup>27</sup>Groseclose and Milyo (2005) report results for six newspapers, nine television broadcasts, one radio broadcast, three national magazines, and one online news source.

$r_z$ :

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

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:

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

Here  $\varepsilon_{zn}$  is a zipcode-specific utility shock,  $\xi_{izn}$  is a household-specific utility shock, and we expect  $\gamma \geq 0$ .<sup>28</sup> Note that we are absorbing newspaper characteristics such as price or quality that do not interact with  $\tilde{y}_z$  into the error term  $\varepsilon_{zn}$ . These characteristics will be absorbed by the market-newspaper fixed effects that we introduce below.

We normalize the utility of the household’s outside option—the consumption bundle that would be chosen conditional on not reading any newspaper—to be 0.<sup>29</sup> 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

---

<sup>28</sup>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. In appendix B, we present an alternative model that allows for within-zipcode heterogeneity in political ideology. Estimates of this model are consistent with our central conclusions regarding the impact of slant on newspaper demand.

<sup>29</sup>Because, for the purposes of estimation, we will include market-newspaper fixed effects in the model, we will effectively allow this outside option to vary nonparametrically by market. The important content of this assumption is then that we do not allow the utility of the outside option to vary across zipcodes. A correlation between the utility of the outside option and zipcode ideology would complicate estimation of our complete structural model of demand, although it would not compromise our more reduced-form finding that slant and zipcode ideology interact positively in determining demand.

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 a standard discrete-choice framework that would require papers to be perfect substitutes at the individual level.<sup>30</sup>

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  $-\gamma(y_n - \tilde{y}_z)^2 + \varepsilon_{zn}$ . This implies that the share of households reading newspaper  $n$  in zipcode  $z$  will be a linear function:

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

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 regarding the determinants of consumer demand and the implications for firms’ optimal choice of slant.

## 4.2 Identification and Estimation

There are two related sources of variation one could 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

---

<sup>30</sup>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 the determinants of consumer demand, and the implications for firms’ profit-maximizing choice of slant, remain similar in this alternative specification.

parameters  $\alpha$ ,  $\beta$ , and  $\gamma$ .

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

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

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 variation across newspapers for a given zipcode. Another is to include newspaper fixed effects and so use only the 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 wish to emphasize that our estimates will depend both on the quality of these controls and on the accuracy of our functional form assumptions. In particular, the quadratic specification of the utility function is what allows us to identify the key parameters while using only variation across zipcodes—a specification with newspaper fixed effects would not be identified with a more flexible interaction between slant ( $y_n$ ) and consumer ideal point ( $\tilde{y}_z$ ). Because our model is built on strong assumptions, we will distinguish two sets of results: results supporting the basic prediction that conservative newspapers circulate relatively more in conservative zipcodes (which do not depend critically on these assumptions), and estimates of the structural parameters (which do).

We will estimate the following model:

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

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  $y_n$ , 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, in the spirit of George and Waldfogel (2003), whether newspapers from highly Republican markets have circulation that is relatively higher in Republican zipcodes within a given 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 endogenous effect of demand patterns on 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.<sup>31</sup> 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.

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 sometimes small. Although we have no natural 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

$$r_z = r_z^* + \frac{r_z^*(1 - r_z^*)}{T_z} v_z \quad (10)$$

where  $T_z$  is the total number of contributions in zipcode  $z$  and  $v_z$  is a standard normal disturbance.<sup>32</sup> 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

---

<sup>31</sup>An endogeneity concern would normally arise in a regression of demand on a product characteristic  $y_n$  chosen optimally by a firm, 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.

<sup>32</sup>Additional evidence corroborates the basic assumptions of the model in equation (10). When we regress our donation-based measure on our registration-based measure for California zipcodes, weighting by the number of donors, the regression constant is economically close to zero (constant = .03, SE = 0.01) and the regression coefficient is economically close to unity (coefficient = 1.04, SE = 0.02). In other words, the data suggest that it is reasonable to approximate our donation-based measure as a noisy measure of the true Republican share. Further estimates indicate that the variance of the error in our donation-based measure scales inversely with the number of donors, in a manner quantitatively consistent with the predictions of equation (10).



regression calibration (Fuller 1987).<sup>33</sup> We would expect this adjustment to increase the magnitude of our coefficients by eliminating attenuation bias.

### 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  $r_z = (y_n - \alpha) / \beta$  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 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.

A series of reduced-form regressions show that the positive interaction term  $y_n r_z$  is extremely robust (results not shown). It survives controls for zipcode characteristics ( $X_z$ ) and zipcode-market interactions  $W_{zn}$ , as well as controls for both (i) interactions between slant ( $y_n$ ) and a full vector of zipcode demographics and (ii) interactions be-

---

<sup>33</sup>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.

tween zipcode republicanism ( $r_z$ ) and a full vector of newspaper market demographics. The key interaction term is also positive and significant in a specification including *zipcode* fixed effects. (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. But the coefficient on the interaction term  $y_n r_z$  is identified, because we observe many zipcodes in which multiple newspapers circulate.) Finally, we find that the interaction term is positive and significant in a specification in which we instrument for  $y_n r_z$  with interactions between  $y_n$  and a vector of zipcode demographics, which serves to mitigate concerns about reverse causality from newspaper slant to zipcode ideology or donations.

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 direct 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 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.<sup>34</sup> 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  $\alpha$  and  $\beta$ .

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.

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 0.01 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 about 3 percent. This drop in demand is equivalent to a reduction of 8 percentage points in the share of households in the zipcode with a college education. Using the own-price elasticity of the *Washington Post* estimated by Gentzkow (forthcoming) as a benchmark, it is also equivalent to the effect of an 8 percent increase in price.<sup>35</sup>

---

<sup>34</sup>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). To verify that the change in coefficients between columns (1) and (2) of table 9 is consistent with a measurement error justification for our IV strategy, we have conducted a simulation in which we assume that the measurement error in slant among newspapers is the same (as a share of the total variance of slant) as among congresspeople. The simulation yields expected OLS and IV coefficients very close to those we report in table 2.

<sup>35</sup>Gentzkow (forthcoming) estimates the own-price elasticity of the *Washington Post* to be .37.

## 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. In particular, assuming a constant per-reader markup across all zipcodes,<sup>36</sup> our demand model straightforwardly implies that choice of slant that maximizes profits in the newspaper’s primary market  $n$ , which we denote  $ideal_n$ , can be written as

$$ideal_n = \alpha + \beta \bar{r}_n \tag{11}$$

where  $\bar{r}_n$  is the average share donating to Republican candidates in the market as a whole.<sup>37</sup> Note that equation (11) is simply the expression for an individual zipcode’s ideal point, with the zipcode level contribution share  $r_z$  replaced by the market average  $\bar{r}_n$ .

Finally, recall that, at the market (rather than zipcode) level, we have data on the Republican vote share  $R_n$ , a less noisy proxy for market ideology than the average donation share  $\bar{r}_n$ .<sup>38</sup> To take advantage of the superior accuracy of this measure, we begin by observing that

$$E(ideal_n | R_n) = \alpha + \beta E(\bar{r}_n | R_n). \tag{12}$$

---

Note that an elasticity less than one makes sense in this case because the paper gets advertising revenue from each marginal reader it gains.

<sup>36</sup>Note that this amounts to assuming that newspapers maximize circulation. In appendix B, we show that allowing advertising revenues per reader to vary across zipcodes as a function of demographic characteristics produces virtually identical conclusions regarding the profit-maximizing level of slant.

<sup>37</sup>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 zipcodes in which the newspaper circulates, use of this more expansive market definition produces an ideal-point measure highly correlated with  $ideal_n$ . While we have computed our slant measure for a total of 433 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.

<sup>38</sup>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.

Modeling the conditional expectation  $E(\bar{r}_n | R_n)$  as linear, it is then straightforward to estimate its parameters, and, with a slight abuse of notation, to write

$$ideal_n = \tilde{\alpha} + \tilde{\beta}R_n \tag{13}$$

where  $\tilde{\alpha}$  and  $\tilde{\beta}$  are the parameters of our demand model, scaled so that they predict  $ideal_n$  as a function of  $R_n$  rather than  $\bar{r}_n$ . We estimate that  $\tilde{\alpha} = 0.41$  and  $\tilde{\beta} = 0.10$ . Because  $\bar{r}_n$  is tightly related to  $R_n$ , these estimates are very similar to the point estimates of  $\alpha$  and  $\beta$ , respectively, from our demand model.

Note that our estimates strongly reject the null hypothesis that  $\tilde{\alpha} = 0$  and  $\tilde{\beta} = 1$ , i.e. that  $ideal_n = R_n$  ( $p < 0.001$ ). Our demand model thus implies that more right-wing markets desire more right-slanted news, but does not predict that a market's preferred newspaper content will be identical to the language of its congressperson. There are several plausible explanations for this pattern. First, the baseline frequency of certain phrases in newspapers could differ mechanically from their baseline frequency in the Congressional Record, for example because the former is written language while the latter is a transcript of spoken language. Second, the language of congresspeople may be subject to higher-frequency political trends—such as the Republican takeover of Congress in 1994—than the language of newspapers.<sup>39</sup> Finally, shocks to either the popularity of particular topics in the news or the relevance of particular topics to legislation could cause random fluctuations in the language of the average newspaper and the average congressperson respectively.

## 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. 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

---

<sup>39</sup>We thank Jim Snyder for pointing out this possibility.

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).<sup>40</sup>

To capture non-pecuniary motives, we assume that each firm  $g$  has an ideal slant  $\mu_g$ , to which 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  $\sigma_\mu^2$ . We assume that owners suffer a quadratic loss for each newspaper that deviates from their ideal point, and that this loss is proportional to the number of households in the newspaper’s market. This is 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.<sup>41</sup>

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

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

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 demand 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

---

<sup>40</sup>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.

<sup>41</sup>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).

solving yields an expression for the firm’s optimal choice of slant  $y_n^*$ :

$$y_n^* = \frac{m_n\gamma}{m_n\gamma + \lambda} ideal_n + \frac{\lambda}{m_n\gamma + \lambda} \mu_g \quad (15)$$

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  $y_n^*$  plus a noise term  $\zeta_n$ . Substituting for the mean of the owners’ tastes  $\bar{\mu}$ , and for the consumer ideal point  $ideal_n$ , yields a random effects model of slant:

$$\hat{y}_n = \frac{\lambda}{m_n\gamma + \lambda} \bar{\mu} + \frac{m_n\gamma}{m_n\gamma + \lambda} (\tilde{\alpha} + \tilde{\beta}R_n) + \frac{\lambda}{m_n\gamma + \lambda} \tilde{\mu}_g + \zeta_n. \quad (16)$$

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

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  $R_n$  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 have a coefficient less than  $\tilde{\beta}$  when regressed on  $R_n$ . In practice, however, because of the uncertainty in our demand-based estimate of  $\tilde{\beta}$ , the first approach is far more powerful, and we therefore focus our study of owner effects on the question of whether the residuals of slant are correlated across papers with the same owner.

We wish to emphasize that our analysis of the firm’s positioning problem is especially simple because we treat each firm as a monopolist. This allows us to analyze a single-agent decision problem rather than a game, and to set aside incentives to change a product’s position in order to soften price competition. The necessary assumption is strong, and it abstracts from some important competitive interactions.

However, we believe it is justified as a first approximation, given that only a handful of papers in our sample face a head-to-head competitor.<sup>42</sup>

## 5.2 Do Newspapers Deviate from Profit Maximization on Average?

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 429 newspapers is 0.47; noting that the average value of  $R_n$  in our sample is 0.53, the average profit-maximizing slant is  $\tilde{\alpha} + \tilde{\beta}(0.53) = 0.46$ . 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.<sup>43</sup> The evidence does not, therefore, support the view that any of the aggregate forces mentioned above are causing large distortions in slant.

Our results do, however, echo Groseclose and Milyo's (2005) finding that average news content resembles a left-of-center congressperson, in that we find that average slant is below the average percent Republican in our sample.<sup>44</sup> However, we find that

---

<sup>42</sup>In appendix B we show that our results on the empirical determinants of slant are robust to excluding data from the small number of cities with competing papers.

<sup>43</sup>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 underlie our calculation of the profit-maximizing level of slant. The confidence interval displayed in figure 4 takes account of both sources of uncertainty. This confidence interval is fairly tight, despite the relative imprecision of our demand estimates, because (as is common in regression models) realizations of the intercept  $\alpha$  are negatively correlated with realizations of the slope  $\beta$  in the demand model.

<sup>44</sup>Indeed, 68 percent of the newspapers in our sample have a slant below the share voting Republican in their primary markets, and a paired  $t$ -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 about



the average profit-maximizing point is *also* to the left of the average congressperson. As we note in subsection 4.4 above, our demand model implies that more right-wing markets desire more right-slanted news, but does not predict that a market’s preferred newspaper content will be identical to the language of its congressperson.

### 5.3 Consumer Characteristics and Slant

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.

In table 3, we examine the robustness of the relationship between slant and consumer characteristics in a regression framework. The first column presents results from an OLS regression of slant on the Republican vote share. The positive relationship observed in figure 5 is highly statistically significant ( $p < 0.001$ ), and variation in consumer preferences explains nearly 20 percent of the variation in slant in this specification.

This first column also allows us to compare the relationship between slant and percent Republican to the one we would expect if news firms were concerned only with maximizing profits from circulation. In particular, as the first panel of the table indicates, if newspapers were pure profit-maximizers a regression of slant on percent Republican  $R_n$  would yield a coefficient equal to  $\tilde{\beta}$ , the effect of a change in  $R_n$  on the profit-maximizing slant  $ideal_n$ . Our point estimates imply that slant responds, if anything, more than expected to the tastes of a newspaper’s customers. However, the actual and expected coefficients are on a similar order of magnitude, and cannot be distinguished statistically once we account for uncertainty in the demand model parameter  $\tilde{\beta}$ .<sup>45</sup> Thus, we cannot reject the pure profit-maximization model, although

---

1.5 standard deviations of our slant measure.

<sup>45</sup>Constraining the coefficient on percent Republican to equal  $\tilde{\beta}$  produces an  $R^2$  of 14, about three-quarters as large as the unconstrained model.

we stress that this test has limited power due to the uncertainty in the demand model parameters.

Columns (2) and (3) of table 3 add owner-specific fixed effects and state fixed effects respectively. Both the coefficient and standard error change only slightly, showing that the relationship is not driven by either owners who purchase papers in similar markets or spurious geographic correlation in slant. Column (4) 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). Although quality does appear to be negatively correlated with our measure of slant, and the magnitude of the coefficient on percent Republican falls somewhat when these controls are added, the relationship remains strong and highly significant.<sup>46</sup>

Results in the next two columns of table 3 address the possibility of reverse causality. Although we have been interpreting the coefficients on percent Republican 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; DellaVigna and Kaplan 2007; Gerber, Karlan, and Bergan 2006). In column (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.<sup>47</sup> In column (6) we instrument for slant with a vector of other pre-determined characteristics of the newspaper's market: log population, percent black, percent with a college degree, percent urban, and log income per capita, all of which are strong predictors of the Republican vote share. Both IV specifications show a strong and statistically significant effect of Republican vote shares on slant, with coefficients that

---

<sup>46</sup>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 B, 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.

<sup>47</sup>In a regression using data from the 421 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 ( $p < 0.001$ ).

are somewhat larger than in our baseline OLS specification.

In column (7) of table 3, we ask whether changes in consumer preferences are associated with changes in newspaper slant, as measured by a preliminary version of our slant measure for the years 2000 and 2004.<sup>48</sup> We regress slant on percent Republican for the years 2000 and 2004, including newspaper fixed effects. We find that changes in slant and changes in vote shares are strongly correlated, and the relationship is both economically large (larger, in fact, than the relationship we estimate in the cross section) and statistically significant ( $p < 0.001$ ). Of course, this test may well be confounded by reverse causality, and should be taken with caution. As a first step towards addressing that concern, we take advantage of the widely noted fact that college-educated voters moved (in relative terms) away from Bush between 2000 and 2004 (Teixeira, 2005), and instrument for the change in percent Republican between 2000 and 2004 with the percent college educated in 2000 (results not shown). This instrumental variables specification also shows a strong, positive, and statistically significant effect of changes in consumer preferences on changes in slant.

## 5.4 Ownership and Slant

We now turn to one of the central questions that motivated this paper—whether the identity of a newspaper’s owner has a significant impact on slant. To address this, we exploit the fact that many newspapers in our sample share the same owner. Of our 429 newspapers, 346 are jointly owned with another paper in our sample. The groups with the largest number of newspapers in our sample are Gannett (91 papers), Media News (40 papers), and Knight Ridder (27 papers). Importantly, many groups are heterogeneous both geographically and politically. For example, the markets where the *New York Times* company owns newspapers range from New York City (28 percent Republican), to Sarasota, FL (54 percent Republican), to Spartanburg, SC (67 percent Republican). The within-group standard deviation of percent voting Republican is .10, only slightly lower than the overall standard deviation of percent voting Republican across all newspaper markets (.12). In half of the groups with

---

<sup>48</sup>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 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 comparability, we standardize the measure in each year to have the same mean and standard deviation as our 2005 measure.

multiple papers in our sample, those papers span more than one state.

Equation (15) predicts that if ownership is important we should expect to see the slant of jointly owned papers pulled away from the profit-maximizing point  $ideal_n$  toward the owner's own ideal point  $\mu_g$ . Two jointly owned papers will then be more similar to each other than two randomly-drawn papers. To test directly for this pattern in the raw data, panel A of figure 6 plots each newspaper's slant against the average slant of other newspapers with the same owner. The graph shows a statistically significant positive correlation ( $p < .0001$ ), providing some preliminary evidence consistent with ownership effects.

There are at least two reasons why we would want to be cautious about drawing conclusions from this picture, however. First, it does not adjust for variation in consumer characteristics. If owners are more likely to own papers in markets with similar political views, the response to consumer characteristics documented in the last section would generate a positive correlation. Second, ownership groups are known to be highly clustered geographically (Lacy and Simon 1997; Martin 2003), and our slant measure is also positively correlated across papers within a state. This latter correlation remains even after adjusting for political characteristics of markets, in part because the kinds of political events whose coverage is picked up by our measure often occur at the state level. Geographic clustering could thus be another source of the positive correlation in figure 6.

Panel B of figure 6 shows the same relationship after adjusting for political characteristics of consumers and geographic clustering. The panel plots the residual from a regression of slant on percent Republican in a paper's market and state fixed effects against the average of this residual among other papers with the same owner. The correlation seen in the previous figure disappears, and the relationship between the variables is no longer significant ( $p = .339$ ).

These findings suggest that owners exert at most a small influence on newspaper slant, once the tendency of owners to cluster in similar markets 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), including the owner-specific random effect  $\tilde{\mu}_g$ . 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 figure 6 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 consumer preferences. 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 0.015. This standard deviation is statistically different from zero and economically nontrivial, accounting for 13 percent of the overall variation in measured slant. These results confirm the correlation visible in panel A of figure 6.

Column (2) presents estimates of equation (16), including both percent Republican and owner-specific random effects. When owner and consumer characteristics are combined, the share of variation explained by ownership falls from 13 percent to 10 percent. This suggests that some of the ownership effect in specification (1) reflects correlation between the distribution of ownership groups and consumer characteristics rather than a causal effect of ownership. In this specification, the share of variation explained by consumer preferences is 0.19, consistent with our earlier findings in table 3.

In columns (3) and (4) of table 4, we add Census division and state fixed effects, respectively. When we control for Census division, a relatively coarse measure of geography, we find that ownership explains a statistically insignificant 6 percent of the within-division slant. 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.

To summarize, our evidence suggests 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).

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 60 percent of the true variation in slant.

One way to strengthen these results would be to use panel data to ask how *changes* in slant and ownership are related. Unfortunately, the limited coverage of news databases for earlier years limits the power of this approach. 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.

If ownership effects are important, we would expect the difference between the mean slant of the acquired and acquiring papers to 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.

A second way to extend our ownership results is to ask whether slant responds more to owner identity in circumstances in which the political returns to changing consumers' voting behavior are large. If ownership effects exist and are driven by a desire to influence elections, this could provide a more powerful way to identify them. To implement this test, we use our preliminary measure of slant in 2004, and test whether slant is more responsive to owner identity (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.<sup>49</sup>

A final way to look for an effect of ownership is to use direct evidence on the political beliefs of owners themselves. Even if the overall impact of slant in our sample is small, it could be the case that a subset of owners with strong political views have a larger effect. As a rough proxy for the ideology of the individuals who control newspaper firms, we have computed the share of donations given by top newspaper executives, and media firms themselves, to Republican and Democratic campaigns.

In figure 7, we plot the relationship between slant and the share of contributions going to Republican candidates for three categories of contributions: (i) those from executives at firms that own multiple U.S. newspapers; (ii) those from executives at independent newspapers (not jointly owned with any other U.S. paper); and (iii) corporate contributions by newspaper firms. The correlation between slant and contributions is weak and insignificant, both overall and taking each of these three groups separately. This remains true in regressions controlling for the percent voting Republican in each paper's market. The results are thus consistent with the view that the identity of a newspaper's owner is not an important determinant of slant.

---

<sup>49</sup>We also find no evidence of significant differences in owner effects between public and private media firms, or between papers in markets with high and low advertising rates. We have also specified and estimated a model in which media firms economize on fixed costs in the production of news by minimizing the diversity of slant across their newspapers. This model predicts that a newspaper's slant will respond to variation in the political attitudes of markets served by co-owned papers. We find no evidence for such an effect.

## 6 Additional Determinants of Newspaper Slant

In this section, we use our data to test for several other influences on newspaper slant that have appeared in prominent economic models of media content.

### 6.1 Pressure from Incumbent Politicians

Besley and Prat (2006) present a model in which pressure or bribes from incumbent politicians causes media firms to produce more favorable coverage than they otherwise would. Though this force is incontrovertibly important in some settings,<sup>50</sup> it is not as obvious whether pressure from incumbent politicians currently affects the slant of coverage by U.S. newspapers.<sup>51</sup>

The evidence discussed in section 5.3 on the average level of slant is informative about the extent to which political pressure shapes coverage at the national level. During the period of our sample, both the White House and Congress were controlled by Republicans. If national-level incumbents were using their power to substantially change news coverage, we would expect to see the average newspaper in our sample pulled to the right of the profit-maximizing point. Although the average paper is slightly to the right, the difference is small and statistically insignificant, suggesting that distortions caused through this channel are not large. Of course this does not rule out the possibility that politicians exert greater influence on specific issues or in certain outlets.

A different possibility is that local politicians influence slant. As a test of this hypothesis, 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 (i.e. leftward shift) in slant of about 0.9 percentage points, with a confidence interval that rules out positive effects larger than about 0.5 percentage points (one-eighth of

---

<sup>50</sup>McMillan and Zoido (2004) present detailed evidence on the bribes paid by the chief of Peru's secret police to media firms in exchange for supportive coverage. The U.S. government paid to place favorable news articles in Iraqi newspapers (Mazzetti and Daragahi 2005). And direct censorship by totalitarian regimes is of course a common occurrence.

<sup>51</sup>Although bribery and direct censorship are rare, it is frequently alleged that U.S. politicians exert influence in more subtle ways, such as cutting off privileged access to government sources for media outlets whose coverage is unfavorable (Ritea 2004).



a standard deviation). We have also 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 0.004, 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 U.S. newspapers’ political positioning.

## 6.2 Tastes of Reporters and Editors

A different possibility is that the personal views of reporters and editors affect slant. 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.

As with influence by politicians, the evidence in section 5.3 is informative about the importance of reporter or editor tastes at the national level. If the composition of the national labor market is such that it is more expensive to hire reporters who will write with a conservative slant, for example, we would expect to see the average paper in our sample pulled to the left of the profit-maximizing point. The fact that this is not the case provides some evidence that employees’ tastes are not decisive in the aggregate.

It could still be the case, however, that variation across markets is driven by the tastes of *local* reporters and editors. This could cloud the interpretation of our earlier estimates if these tastes vary in a similar way to the tastes of consumers, as it would induce correlation between actual slant and market-level politics.

Although we cannot estimate the importance of workers’ tastes directly, we believe for three 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, mobility across markets must be 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).<sup>52</sup> 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.<sup>53</sup> 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 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 would cost the firm *more* than 3 percent of variable profits. To get a sense of the wage effect this would imply, calculations based on data in Gentzkow (forthcoming) suggest

---

<sup>52</sup>We are extremely grateful to Lisa Kahn for providing the appropriate calculations from the NLSY.

<sup>53</sup>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.

that the *Washington Post*'s variable profit in 2004 was on the order of \$500 million.<sup>54</sup> *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 \$90,000 per year and the average editor's salary is \$125,000 per year (probably overestimates), we estimate the *Post*'s wage bill for reporters and editors to be about \$42 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  $(500/42)*3.5$  percent = 36 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.

Finally, we have computed an alternate version of our slant measure using only stories written by newspapers' Washington D.C. bureaus. Because the reporters and editors in these bureaus live and work in Washington D.C., barriers to mobility would not induce a correlation between these reporters' views and those of a paper's local market. Although many papers do not have Washington bureaus, and our power is correspondingly lower, a regression of the slant of these stories on the percent Republican in the newspaper's market yields a positive and statistically significant coefficient, with a value not statistically distinguishable from the coefficient we obtain when we use the overall slant measure.

### 6.3 Newspaper Competition

Mullainathan and Shleifer's (2005) model predicts that, because product differentiation softens price competition, newspapers in competitive markets will be more "biased" than monopoly papers, in the sense that competitive papers' slants will deviate from one another and from the preferences of the median consumer. Our data contain only very limited information about the role of competition, as our sample includes only 29 newspapers in cities with two or more independently owned, competing newspapers. However, as a crude test of Mullainathan and Shleifer's (2005)

---

<sup>54</sup>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.

hypothesis we can ask whether newspapers in multi-paper cities display a greater deviation from the preferences of the average consumer than do newspapers in one-paper cities. To implement this test, we regress slant on the Republican share of the two-party vote in 2004 among newspapers in single-paper cities, and compute the predicted value as a measure of the average consumer’s preferred slant. We then compute the difference between predicted slant and the observed slant of each paper. We find that the average squared difference is higher in multi-paper cities than in single-paper cities, though the difference is not statistically significant. Thus, our evidence is directionally consistent with the hypothesis that competition between newspapers creates greater heterogeneity in slant. It is important to stress that, with so few observations in multi-paper cities, the power of this test is limited.

## 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 roughly 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. 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. We also find no evidence that pressure from incumbent politicians or the tastes of reporters are important drivers of slant.

Taken together, our findings suggest that ownership diversity may not be a critical

precondition for ideological diversity in the media, at least along the dimension we consider. This conclusion has broad implications for the regulation of ownership in the media. We note, however, that our results may well be different in settings with significantly different legal or institutional environments—less developed markets, more state ownership, less freedom of the press—and that determining how these factors affect the determination of slant remains an open question.

## A Appendix: Details on News Searches

### A.1 Mechanics of Searches

Following the steps outlined in 3.1, we identify 1,000 phrases to use in our analysis. We wish to count the number of times each of these 1,000 phrases appears in each of our 433 newspapers, using the ProQuest and NewsLibrary databases. As discussed in the text, we use counts from the NewsLibrary database when available and counts from the ProQuest database otherwise.

Recall that we apply two pre-processing steps to the Congressional Record as described in section 2. First, we drop extremely common words such as “to,” “from,” and “the,” using a standard list of these “stopwords” from Fox (1990). Second, we run the text through an algorithm called the “Porter Stemmer” that strips words down to shared linguistic roots. This means that phrases in the Congressional Record that differ only in either stopwords or suffixes are equivalent in our algorithm. For example, “war on terror,” “war against terror,” and “wars on terror” would all appear in the pre-processed Congressional Record as “war terror” and thus be treated as the same phrase. Each of our 1,000 phrases thus corresponds to a group of one, two, or several “original phrases”, and it is these original phrases that we search for in the databases.

There are two reasons why the set of original phrases we search is slightly reduced. First, the ProQuest database limits search strings to 75 characters. We therefore drop any original phrase longer than 75 characters. Second, our database of Congressional Record text has improved over time as we have adjusted for errors in the source website and improved our parsing algorithm. The set of original phrases included in each group is based on a slightly older version of the Congressional Record text than the one used for our main analysis, and so omits some relatively rare original phrases.

We search for each group of original phrases (connected with the “OR” operator) in the “All Text” field (NewsLibrary) or “Document Text” field (ProQuest), restricted to 2005 and with the following terms excluded from the “Headline” and “Author” fields: “editor,” “editorial,” “associated press,” “ap,” “opinion,” “op-ed,” and “letter.”

### A.2 Audit Study

Our searches are designed to isolate the slant of news content produced independently by each paper. The way stories are archived and classified in the databases means that we can only imperfectly separate these stories from other kinds of content such as opinion pieces and wire stories. To provide a more precise picture of the kinds of content we are measuring, we have audited the results for seven phrases chosen from table 1. For each phrase, we looked at the full set of hits for the papers included in the NewsLibrary database and recorded whether they appeared to be: (i) independently produced news stories; (ii) AP wire stories; (iii) other wire stories; (iv) letters to the editor; (v) opinion pieces (including unsigned editorials). Because we do not have access to the full text of articles in NewsLibrary, this classification is based on the headline and first paragraph of the story.

In a separate exercise, we use results from the papers we can search in the ProQuest

database (for which we can retrieve full text articles) to record the number of times each phrase appears in quotation.

The results are shown in appendix table 1. Overall, approximately 71 percent of our hits are independently-produced news stories. Of the remainder, 23 percent are either clearly or possibly opinion, 3 percent are letters to the editor, and 3 percent are wire stories. The table also shows that these shares are heterogeneous across phrases. For example, the share of opinion pieces ranges from 11 percent for “global war on terrorism” to 50 percent for “death tax.” The results also show that only 10 percent of our hits appear in quotations, with the share ranging from 3 percent for “child support enforcement” to 36 percent for “death tax.” We have also spot checked the articles that are being excluded from our search results and verified that virtually all of them are, as desired, either wire stories or opinion pieces.

Taken together, the results confirm that our measure is primarily picking up the slant of independently-produced news stories, with some weight given to opinion pieces.

## B Appendix: 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 2. 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 percent voting for Bush in 2004 in its market, 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, along with the expected coefficient on percent Republican ( $\hat{\beta}$ ) from our preferred demand model.

In row (2) of appendix table 2, we relax the assumption that newspapers’ per-reader markup is constant across zipcodes, and allow for a more realistic profit function in which readers’ value to advertisers differs according to their demographics. To estimate the advertising value of each zipcode, we use cross-newspaper variation in ad rates to estimate the approximate advertising value of different demographics. We then apply the coefficients from this regression to each zipcode’s demographic characteristics, to produce an estimate of the advertising value of readers in each zipcode. Finally, we combine these estimates with data on cover prices and estimates of marginal costs (following Gentzkow forthcoming) to generate a per-reader markup for each zipcode. As the table shows, computing the profit-maximizing slant using this richer model does not meaningfully affect the relationship between slant and market percent Republican implied by our demand model.

In row (3) of appendix table 2, we relax a second assumption of our demand framework, namely that all households in a given zipcode have identical preferences for slant. As an alternative, we assume that each household  $i$  in zipcode  $z$  has an

ideology  $q_{iz}$  and an ideal slant of

$$\tilde{y}_{iz} = \alpha + \beta q_{iz} \quad (17)$$

We then assume that  $q_{iz} = \theta_z + \zeta_{iz}$ , where  $\zeta_{iz}$  is a standard normal random variable with cdf  $\Phi$ , and  $\theta_z$  is a shift parameter governing the mean ideology in a zipcode. Maintaining the demand model’s other assumptions, it is straightforward to show that, in this case, our equation (9) becomes (with slight abuse of notation):

$$S_{zn} = \delta_{mn} + 2\gamma\beta y_n \theta_z - 2\gamma\alpha\beta\theta_z - \gamma\beta^2 (\theta_z^2 + 1) + X_z\phi_1 + W_{zn}\phi_2 + \varepsilon_{zn}. \quad (18)$$

If we further suppose that a household votes for Bush in 2004 if and only if  $\theta_z \geq 0$  (where 0 could be replaced by any arbitrary constant), then  $\theta_z = \Phi^{-1}(r_z)$ . It is then straightforward to estimate model (18), by estimating  $\theta_z$  using empirically observed donation shares, and adjusting for measurement error using the delta method. As the table shows, this alternative demand framework makes similar (but less precise) predictions about the empirical relationship between slant and consumer preferences, relative to our baseline model.

In row (4) of appendix table 2, 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 2, 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 in which we replace 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. Applied to newspapers, 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 2, 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 2, 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, and has different units (and a different sign) from the percent Republican in the congressperson's constituency. Other than the inconsequential change in units, the ADA-based measure does not yield different conclusions.

## References

- Aas, Kjersti, & Eikvil, Line. 1999. Text categorisation: A survey. *Norwegian Computing Center Mimeograph*, June.
- Akhavan-Majid, Roya, Rife, Anita, & Gopinath, Sheila. 1991. Chain ownership and editorial independence: A case study of Gannett newspapers. *Journalism and Mass Communication Quarterly*, **68**(Spring-Summer), 59–66.
- Alterman, Eric. 2003. *What Liberal Media? The Truth about Bias and the News*. Basic Books.
- Ansolabehere, Stephen, Lessem, Rebecca, & Snyder Jr., James M. 2006. The orientation of newspaper endorsements in U.S. elections, 1940-2002. *Quarterly Journal of Political Science*, **1**(4), 393–404.
- Antweiler, Werner, & Frank, Murray Z. 2004. Is all that talk just noise? The information content of Internet message boards. *Journal of Finance*, **59**(3), 1259–1294.
- Bagdikian, Ben H. 2000. *The Media Monopoly*. 6th edn. Boston: Beacon Press.
- Balan, David J., DeGraba, Patrick, & Wickelgren, Abraham L. 2005. Ideological persuasion in the media. *Federal Trade Commission mimeograph*, April.
- Baron, David P. 2006. Persistent media bias. *Journal of Public Economics*, **90**, 1–36.
- Becker, Gary S. 1957. *The Economics of Discrimination*. 2nd (1971) edn. University of Chicago Press.
- Berry, Steven, & Waldfogel, Joel. 2003. Product Quality and Market Size. *NBER Working Paper No. 9675*.
- Besley, Timothy, & Prat, Andrea. 2006. Handcuffs for the grabbing hand? Media capture and government accountability. *American Economic Review*, **96**(3), 720–736.
- Burrelle's Information Services. 2001. *Burrelle's/Luce Media Directory - 2001 Edition*.
- Commission on Freedom of the Press. 1947. *A free and responsible press : a general report on mass communication: newspapers, radio, motion pictures, magazines, and books*. Chicago, IL: The University of Chicago Press.
- Coulter, Ann. 2003. *Slander: Liberal Lies about the American Right*. Three Rivers Press.
- DellaVigna, Stefano, & Kaplan, Ethan. 2007. The Fox News effect: Media bias and voting. *Quarterly Journal of Economics*, August.
- Dranove, David, Gron, Anne, & Mazzeo, Michael J. 2003. Differentiation and competition in HMO markets. *Journal of Industrial Economics*, **51**(4), 433–454.

- Dubé, Jean-Pierre, Hitsch, Günter J., & Manchanda, Puneet. 2005. An empirical model of advertising dynamics. *Quantitative Marketing and Economics*, **3**(2), 107–144.
- Einav, Liran. 2007. Seasonality in the US Motion Picture Industry. *RAND Journal of Economics*. Forthcoming.
- Federal Communications Commission. 2003. *Report and Order and Notice of Proposed Rulemaking*. Washington, DC: Federal Communications Commission.
- Fox, C. 1990. A stop list for general text. *SIGIR FORUM*, **24**(4).
- Franken, Al. 2003. *Lies and the Lying Liars Who Tell Them: A Fair and Balanced Look at the Right*. E P Dutton.
- Fuller, Wayne A. 1987. *Measurement Error Models*. Wiley series in probability and mathematical statistics. New York: Wiley.
- Gentzkow, Matthew. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*. Forthcoming.
- Gentzkow, Matthew. 2006. Television and voter turnout. *Quarterly Journal of Economics*, **121**(3).
- Gentzkow, Matthew, & Shapiro, Jesse M. 2006. Media bias and reputation. *Journal of Political Economy*, **114**(2), 280–316.
- Gentzkow, Matthew A., & Shapiro, Jesse M. 2004. Media, Education, and anti-Americanism in the Muslim World. *Journal of Economic Perspectives*, **18**(3), 117–133.
- Gentzkow, Matthew A., Glaeser, Edward L., & Goldin, Claudia D. 2006. The Rise of the Fourth Estate: How Newspapers Became Informative and Why it Mattered. *Chap. 6, pages 187–230 of: Glaeser, Edward L., & Goldin, Claudia (eds), Corruption and Reform: Lessons from America's Economic History*. Chicago, IL: University of Chicago Press.
- George, Lisa. 2001. What's fit to print: The effect of ownership concentration on product variety in newspaper markets. *Michigan State University Mimeograph*, September.
- George, Lisa, & Waldfogel, Joel. 2003. Who affects whom in daily newspaper markets? *Journal of Political Economy*, **111**(4), 765–784.
- Gerber, Alan, Karlan, Dean, & Bergan, Daniel. 2006. Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. *Yale University mimeograph*, February.

- Gimpel, James G., Lee, Frances E., & Kaminski, Joshua. 2006. The political geography of campaign contributions in American politics. *Journal of Politics*. Forthcoming.
- Glaeser, Edward L., Ponzetto, Giacomo A. M., & Shapiro, Jesse M. 2005. Strategic extremism: Why Republicans and Democrats divide on religious values. *Quarterly Journal of Economics*, **120**(4), 1283–1330.
- Glasser, Theodore L., Allen, David S., & Blanks, S. Elizabeth. 1989. The influence of chain ownership on news play: A case study. *Journalism Quarterly*, **66**(Fall), 607–614.
- Goldberg, Bernard. 2003. *Bias : A CBS Insider Exposes How the Media Distort the News*. Perennial.
- Graetz, Michael J., & Shapiro, Ian. 2005. *Death by a thousand cuts: The fight over taxing inherited wealth*. Princeton, NJ: Princeton University Press.
- Groseclose, Tim, & Milyo, Jeff. 2005. A measure of media bias. *Quarterly Journal of Economics*, **120**(4), 1191–1237.
- Groseclose, Tim, Levitt, Steven D., & Snyder, Jr., James M. 1999. Comparing interest group scores across time and chambers: Adjusted ADA scores for the U.S. Congress. *American Political Science Review*, **93**(1), 33–50.
- Hamilton, James T. 2004. *All the News That's Fit to Sell: How the Market Transforms Information into News*. Princeton, NJ: Princeton University Press.
- Harris Interactive. 2006. Seven in 10 U.S. Adults Say They Watch Broadcast News at Least Several Times a Week. *The Harris Poll*, **20**(February 24).
- Hayakawa, S. I. 1942. *Language in Thought and Action*. 3rd (1972) edn. New York: Harcourt Brace Jovanovich Inc.
- Lacy, Stephen, & Simon, Todd F. 1997. Intercounty group ownership of daily newspapers and the decline of competition for readers. *Journalism and Mass Communication Quarterly*, **74**(4), 814–825.
- Larcinese, Valentino, Puglisi, Riccardo, & Snyder, James M. 2006. Partisan Bias in Economic News: Evidence on the Agenda-Setting Behavior of U.S. Newspapers. *London School of Economics Mimeo*.
- Laver, Michael, Benoit, Kenneth, & Garry, John. 2003. Extracting policy positions from political texts using words as data. *American Political Science Review*, **97**(2), 311–331.
- Luntz, Frank. 2005. *Learning from 2004...winning in 2006*. Luntz Research Companies.

- Martin, Hugh J. 2003. 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.
- Mazzeo, Michael J. 2002a. Product choice and oligopoly market structure. *RAND Journal of Economics*, **33**(2), 221–242.
- Mazzeo, Michael J. 2002b. Competitive outcomes in product-differentiated oligopoly. *Review of Economics and Statistics*, **84**(4), 716–728.
- Mazzetti, Mark, & Daragahi, Borzou. 2005. The Conflict in Iraq; U.S. Military Covertly Pays to Run Stories in Iraqi Press. *Los Angeles Times*, December 18.
- McMillan, John, & Zoido, Pablo. 2004. A measure of media bias. *Journal of Economic Perspectives*, **18**(4), 69–92.
- Mullainathan, Sendhil, & Shleifer, Andrei. 2005. The Market for News. *American Economic Review*, **95**(4), 1031–1053.
- Myers, Caitlin Knowles. 2005. Discrimination as a competitive device: The case of local television news. *IZA Discussion Paper No. 1802*, October.
- Newspaper Association of America. 2006. The Source: Newspapers by the Numbers. <http://www.naa.org/theforce/6.asp>.
- Pritchard, David. 2002. 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.
- Puglisi, Riccardo. 2006. Being the New York Times: the political behavior of a newspaper. *Massachusetts Institute of Technology Working Paper*, March.
- Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
- Ritea, Steve. 2004. Going It Alone. *American Journalism Review*, August/September.
- Ruggles, Steven, Sobek, Matthew, Alexander, Trent, Fitch, Catherine A., Goeken, Ronald, Hall, Patricia Kelly, King, Miriam, & Ronnander, Chad. 2004. *Integrated Public Use Microdata Series: Version 3.0*. Minneapolis, MN: Minnesota Population Center. <http://www.ipums.org>.
- Seim, Katja. 2006. An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics*. Forthcoming.
- Stevenson, Richard W. 2005. President makes it clear: Phrase is "war on terror". *New York Times*, August 4, A.12. Eric Schmitt contributed reporting to this article from Washington.

- Stock, James H., & Yogo, Motohiro. 2002. Testing for weak instruments in linear IV regression. *NBER Technical Working Paper 284*.
- Strömberg, David. 2004. Radio's impact on public spending. *Quarterly Journal of Economics*, **119**(1), 189–221.
- Suen, Wing. 2004. The Self-Perpetuation of Biased Beliefs. *Economic Journal*, **114**(April), 377–396.
- Sweeting, Andrew. 2006. Too much rock and roll? Station ownership, programming, and listenership in the music radio industry. *Northwestern University Mimeograph*, January.
- Teixeira, Ruy. June 8, 2005. Democrats and the middle class. *Public Opinion Watch*.
- US Supreme Court. 1945. *Associated Press v. United States*. US Supreme Court.
- Weaver, David H., & Wilhoit, G. Cleveland. 1996. *The American Journalist in the 1990s: U.S. News People at the End of an Era*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wozniak, Abigail K. 2006. Making Their Own Luck: Educational Differences in the Migration Responses of Young Workers to Local Labor Market Conditions. *University of Notre Dame Mimeograph*, January.

**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

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	gas natural gas
deficit reduction bill	american farm bureau	supply of 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

Notes: Table shows top 150 Democratic and Republican phrases respectively, ranked by  $\chi^2$ . 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) $\times$ Slant	0.1800 (0.0705)	0.6115 (0.1868)	1.0790 (0.3120)	0.7781 (0.2925)	0.8323 (0.3011)
Zip share donating to Republicans	-0.0176 (0.0352)	-0.2158 (0.0873)	-0.4234 (0.1432)	-0.3082 (0.1374)	-0.3310 (0.1409)
(Zip share donating to Republicans) <sup>2</sup>	-0.0615 (0.0079)	-0.0625 (0.0077)	-0.0659 (0.0127)	-0.0395 (0.0122)	-0.0417 (0.0119)
Market-newspaper FE?	X	X	X	X	X
Zipcode demographics?				X	X
Zipcode X market char.?					X
Estimate of $\alpha$ (Confidence interval)	0.0976 (-0.94,0.29)	0.3529 (0.19,0.40)	0.3924 (0.30,0.42)	0.3961 (0.21,0.44)	0.3976 (0.24,0.44)
Estimate of $\beta$ (Confidence interval)	0.6838 (0.34,2.64)	0.2044 (0.12,0.52)	0.1222 (0.06,0.30)	0.1016 (0.03,0.41)	0.1003 (0.03,0.37)
Estimate of $\gamma$ (Confidence interval)	0.1316 (0.01,0.44)	1.4958 (0.23,4.06)	4.4165 (0.82,12.2)	3.8285 (0.23,18.1)	4.1501 (0.32,17.4)
Number of observations	62177	62177	62177	62177	62177
Number of newspapers	297	297	297	297	297

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** *Reduced-form evidence on consumer characteristics and newspaper slant*

		Coefficient on percent Republican predicted by demand model ( $\hat{\beta}$ )		0.1033				
		(Confidence interval)		(0.03, 0.37)				
Dependent variable: Slant index ( $\hat{y}_n$ )		(1)	(2)	(3)	(4)	(5)	(6)	(7)
		OLS	FE	FE	FE	IV	IV	FE
Percent Republican in newspaper's market		0.1460 (0.0148)	0.1603 (0.0191)	0.1717 (0.0165)	0.1111 (0.0144)	0.1605 (0.0612)	0.2559 (0.0263)	0.2877 (0.0873)
Sample		2005	2005	2005	2005	2005	2005	2000 & 2004
Ownership group fixed effects?			X					
State fixed effects?				X				
Control for paper chars?					X			
Newspaper fixed effects?								X
Instruments						% church	log pop, % black % college, % urban log income per cap	
Number of observations		429	429	429	429	421	429	638
$R^2$		0.1859	0.4445	0.5127	0.3071	—	—	0.8213

Source: Authors' calculations based on 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. Newspaper market is defined as the newspaper's primary metropolitan statistical area if available, and the newspaper's county if not. Controls for paper characteristics are log(number of pages), log(number of employees), and number of Pulitzer prizes won since 1970. Newspaper market is defined as the newspaper's primary metropolitan statistical area if available, and the newspaper's county if not.

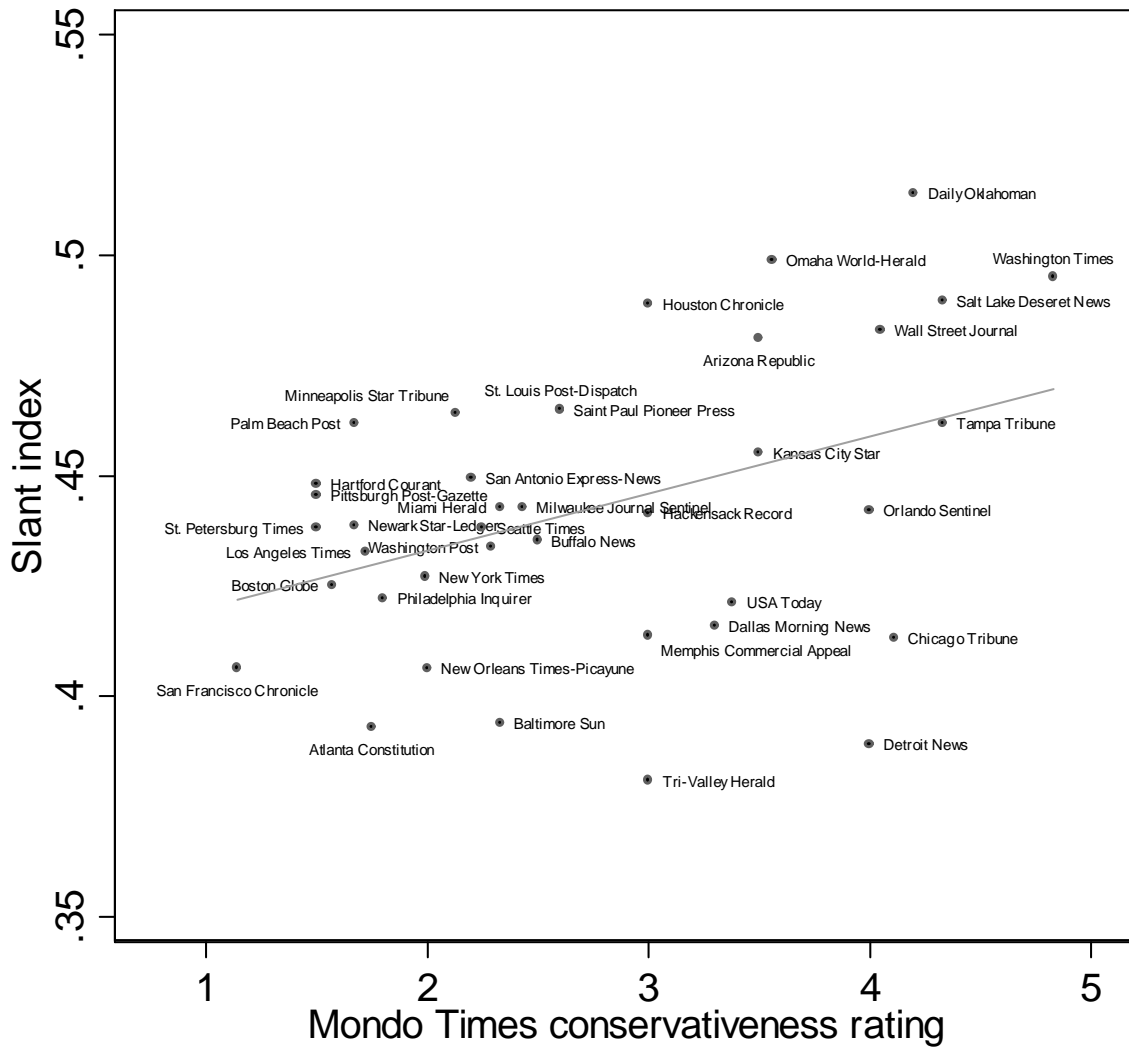
**Table 4** *Decomposing the variation in newspaper slant*

Dependent variable: Slant index ( $\hat{y}_n$ )				
	(1)	(2)	(3)	(4)
Coefficient on percent Republican predicted by demand model ( $\tilde{\beta}$ )				0.1033
(Confidence interval)				(0.03,0.37)
Percent Republican in newspaper's market	—	0.1466 (0.0152)	0.1594 (0.0159)	0.1717 (0.0155)
Census division fixed effects?			X	
State fixed effects?				X
Standard deviation of ownership effect	0.0145 (0.0040)	0.0125 (0.0038)	0.0087 (0.0042)	0.0000 (0.0051)
Ownership share of residual variation	0.1312 (0.0688)	0.0987 (0.0541)	0.0571 (0.0413)	0.0000 (0.0603)
Consumer share of residual variation		0.1875 (0.0389)	0.1923 (0.0382)	0.2225 (0.0402)
Number of observations	429	429	429	429
Number of multi-paper groups	36	36	36	36

Source: Authors' calculations based on 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. 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 percent Republican respectively; in columns (2), (4) and (5) the share(s) are computed after partialling for group, division, and state fixed effects respectively.

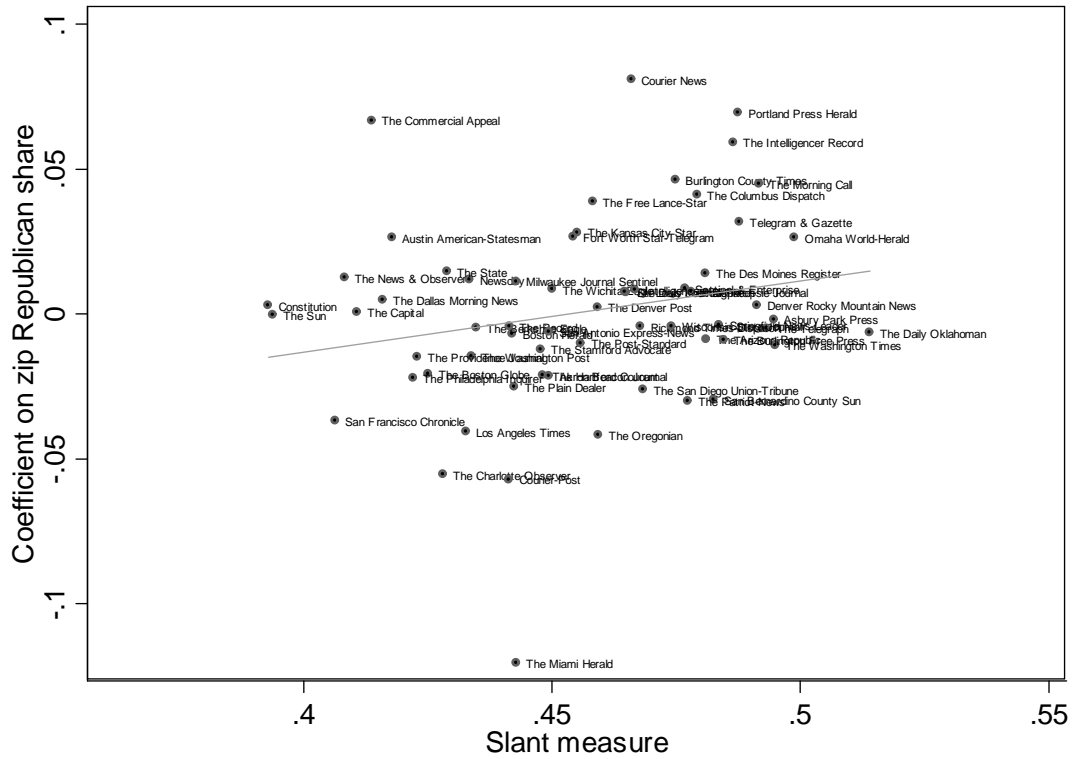
**Figure 1** *Language-based and reader-submitted ratings of slant*



Source: Authors' calculations based on 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.

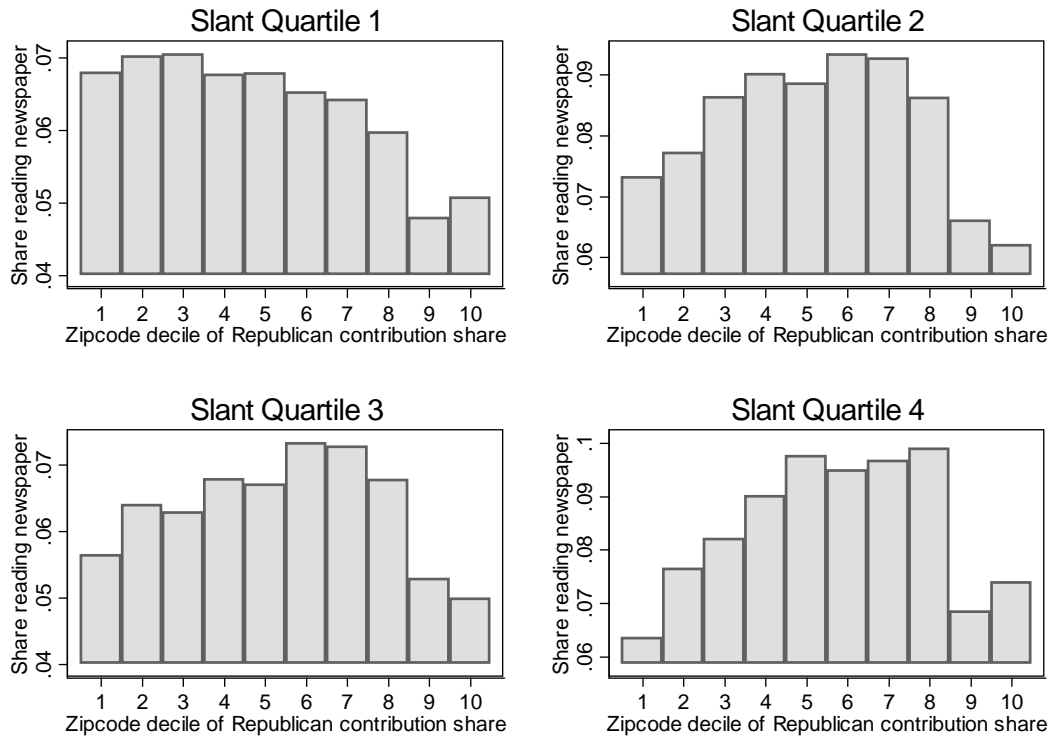
**Figure 2** *Newspaper slant and consumer demand*



Source: Authors' calculations based on Audit Bureau of Circulations (newspaper subscriptions), Federal Election Commission (campaign contributions)

Notes: Y-axis shows the estimated coefficient in a regression of the share of households in the zipcode reading each newspaper on the zipcode share Republican, for newspapers circulating in at least 300 zipcodes. X-axis shows slant measure.

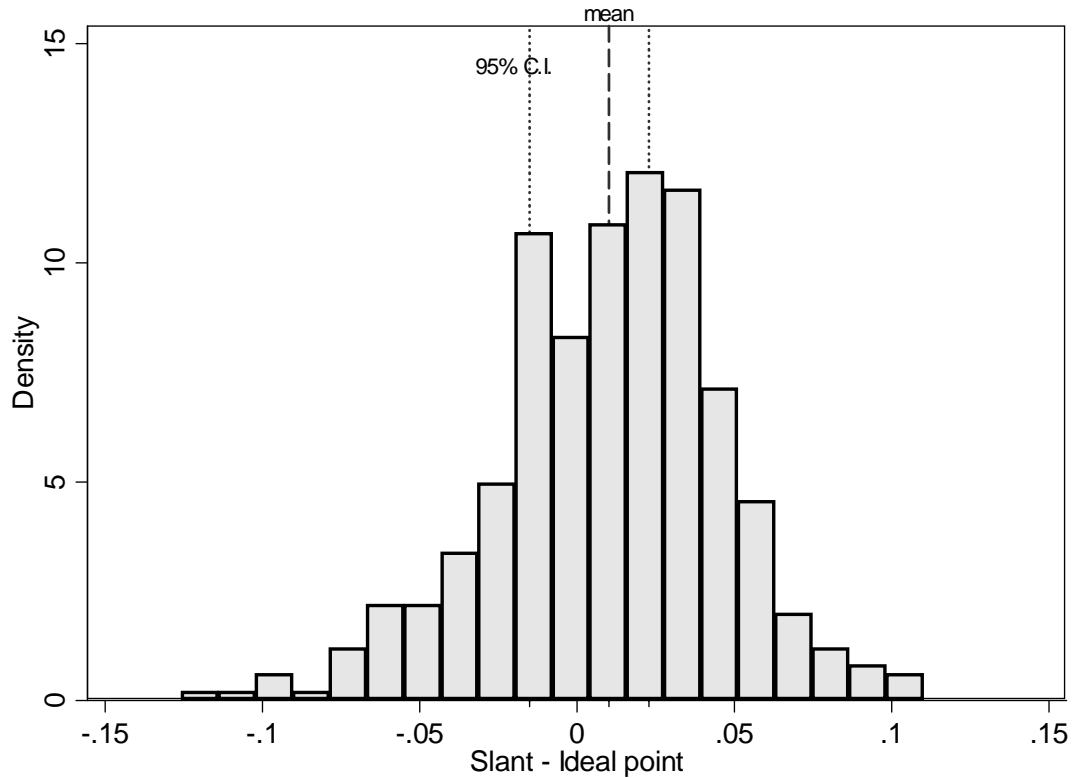
**Figure 3** *Newspaper slant and variation in consumer demand*



Source: Authors' calculations based on 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.

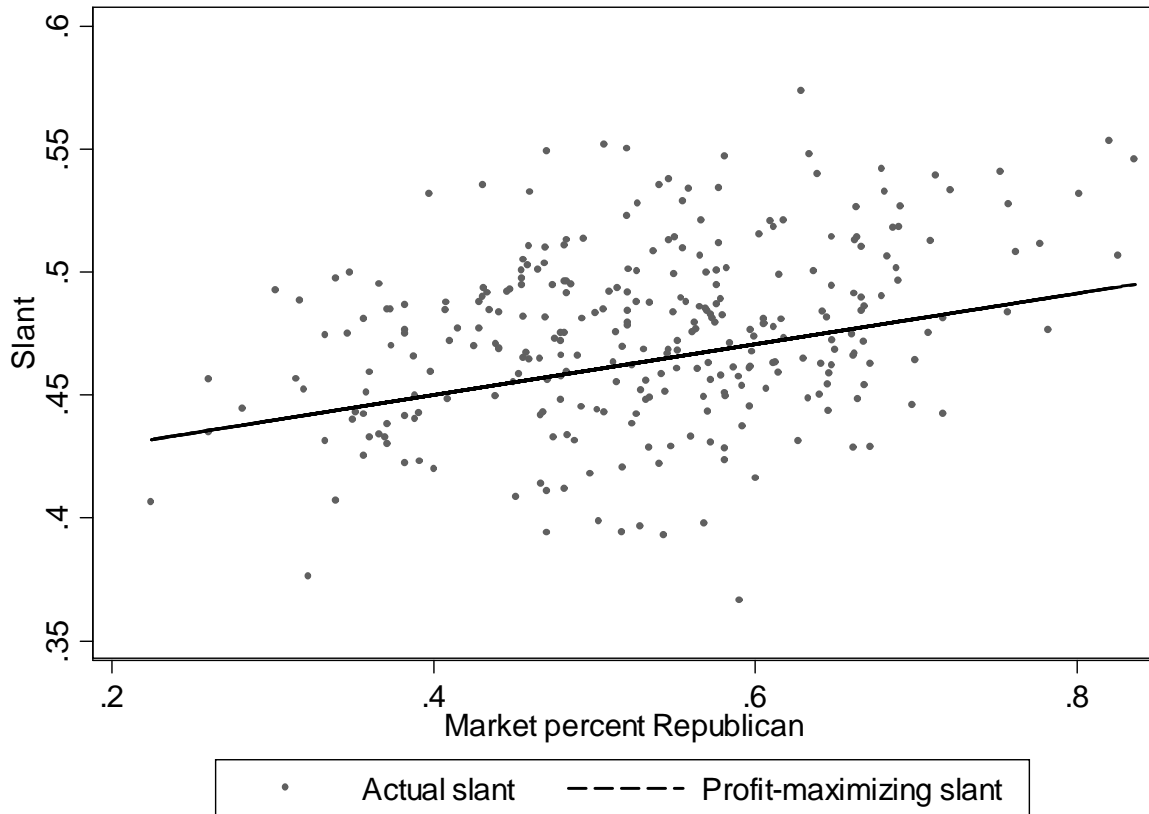
**Figure 4** *Differences between slant and predicted ideal point*



Source: Authors' calculations based on 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*



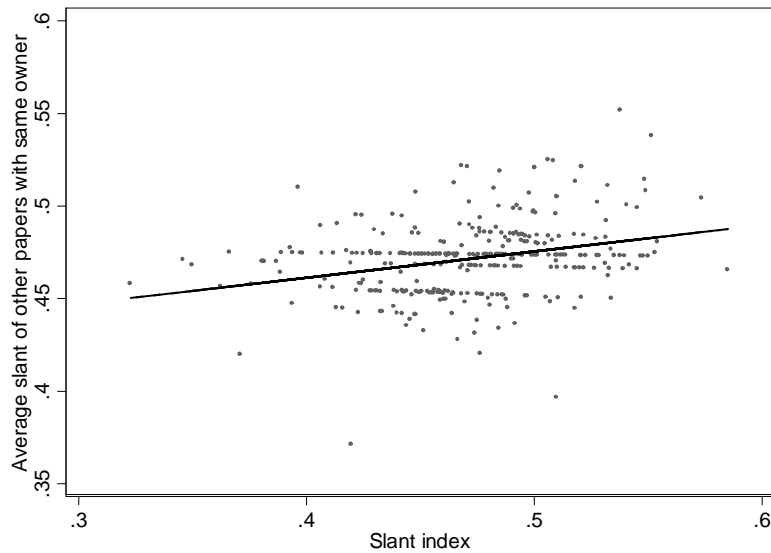
Source: Authors' calculations based on U.S. Presidential Atlas (county-level voting), Editor and Publisher International Yearbook 2000-2005 (newspaper location).

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 slant. Newspaper market is defined as the newspaper's primary metropolitan statistical area if available, and the newspaper's county if not.

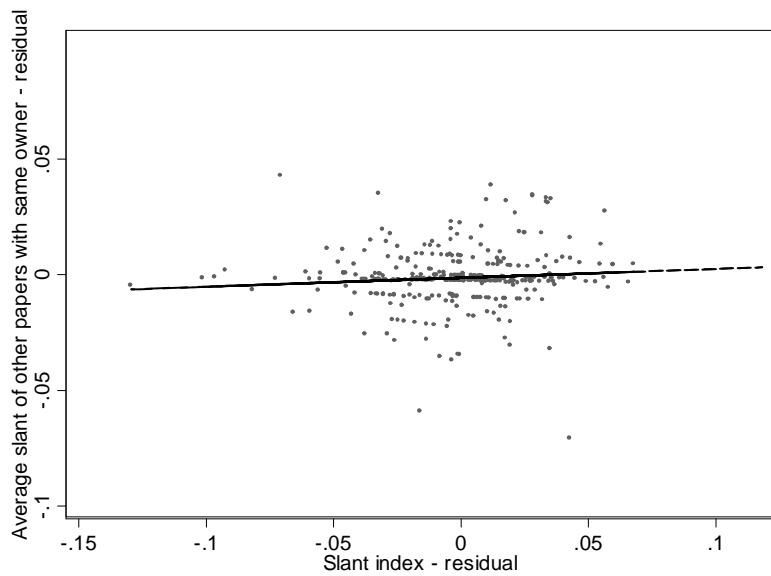


**Figure 6** *Newspaper slant and ownership*

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



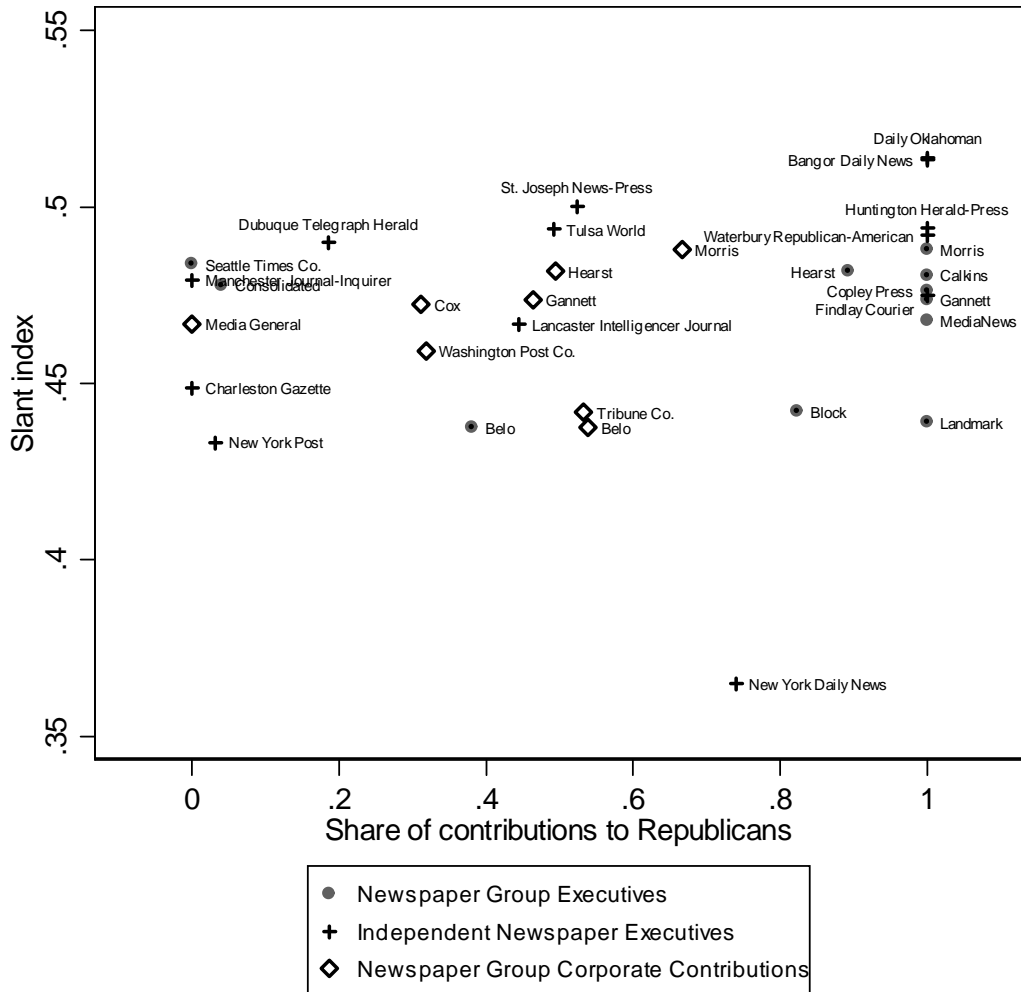
Panel B: Newspaper slant and slant of co-owned papers, controlling for consumer preferences and state



Source: Authors' calculations based on Editor and Publisher International Yearbook 2000-2005 (newspaper location and ownership).

Notes: See section 3 for derivation of slant index. 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 percent Republican in market and dummies for the state in which the newspaper is located.

**Figure 7** *Newspaper slant and political contributions*



Source: Authors' calculations based on Editor and Publisher International Yearbook 2000-2005 (newspaper ownership), Federal Election Commission (donations of executives), Center for Public Integrity (corporate donations).

Notes: Figure shows average slant of newspapers owned by a firm graphed against the share of total dollars going to Republicans within each category of contributions.

**Appendix Table 1** *Audit of search results*

Phrase	Share of hits that are							
	Total hits	Share of hits in quotes	AP wire stories	Other wire stories	Letters to the editor	Maybe opinion	Clearly opinion	Independently-produced news
Global war on terrorism	2064	16%	3%	4%	1%	2%	10%	80%
Malpractice insurance	2190	5%	0%	0%	1%	3%	12%	84%
Universal health care	1523	9%	1%	0%	7%	8%	28%	56%
Assault weapons	1411	9%	3%	12%	4%	1%	25%	56%
Child support enforcement	1054	3%	0%	0%	1%	2%	11%	86%
Public broadcasting	3375	8%	1%	0%	2%	4%	22%	71%
Death tax	595	36%	0%	0%	2%	5%	46%	47%
Average (hit-weighted)		10%	1%	2%	3%	3%	19%	71%

Source: Authors' calculations based on ProQuest and NewsLibrary database searches.

Notes: See appendix A for details.

**Appendix Table 2** *Additional robustness checks*

Specification	Coefficient on percent Republican		Share of residual variation	
	Expected ( $\tilde{\beta}$ )	Actual	Owner	Consumer
(1) Baseline	0.1033 (0.03,0.37)	0.1717 (0.0155)	0.0000 (0.0603)	0.2225 (0.0402)
(2) Weighting zipcodes by predicted profits	0.1044 (0.04,0.38)	—	—	—
(3) Allowing within-zipcode heterogeneity in ideology	0.1788 (-0.85,1.64)	—	—	—
(4) Exclude multi-paper cities from supply model	0.1033 (0.03,0.37)	0.1556 (0.0162)	0.0000 (0.0556)	0.1901 (0.0396)
(5) Controlling for predicted sophistication	0.1033 (0.03,0.37)	0.1741 (0.0156)	0.0000 (0.0601)	0.2249 (0.0403)
(6) Tightening cutoffs on phrase counts by 5%	0.0985 (0.04,0.36)	0.1680 (0.0166)	0.0000 (0.0557)	0.1928 (0.0381)
(7) Measuring ideology with adjusted ADA score	-21.37 (-78.6,-6.92)	-31.31 (3.022)	0.0000 (0.0627)	0.2001 (0.0386)

Source: Authors' calculations based on 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 expected coefficient on percent Republican ( $\tilde{\beta}$ ). 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 percent Republican respectively; these shares are computed after partialling for state fixed effects. See appendix B for details.