Demand for (Un)Biased News: Government Control in Online News Markets

Andrey Simonov  Justin Rao
Columbia University  HomeAway, Inc.

11th October 2017

Consumers in many countries get news from government-controlled sources even when independent sources are available. Do consumers like pro-government bias in the news, quality of the government-controlled outlets, or both? We characterize government control in the Russian online news market use publication records, and combine it with the click-level browsing data to estimate the demand for news. Government-controlled outlets would have 20.3% higher market share if they reported sensitive news like the independent outlets and 44.6% lower market share if they had the quality of the independent outlets. The higher quality of the government-controlled outlets substantially increase their media power.

*Simonov: asimonov@gsb.columbia.edu; Rao: justinmrao@outlook.com. Andrey is grateful to his advisors and dissertation committee members Pradeep Chintagunta, Jean-Pierre Dubé, Matthew Gentzkow and Günter Hitsch for their guidance and support. We thank Matt Goldman, Ali Hortaçsu, Sanjog Misra, Chris Nosko, Bradley Shapiro, Raluca Ursu, Elisabeth Honka, Ali Yurukoglu, Ruben Enikolopov, participants of the 16th Trans-Atlantic Doctoral Conference, 2016 AMA-Sheth Doctoral Consortium, 2016 Economics of Media and Communications Conference, 2017 Marketing Science Conference and the 15th Media Economics Workshop, seminar participants at the University of Chicago, Stigler Center at Booth School of Business, HKUST, University of Minnesota, University of Rochester, UCSD, Duke, LBS, Columbia, Southern Methodist University, Berkeley, CMU, UCLA and University of Southern California for advice and suggestions. All opinions represent our own and not those of our current or past employers. All remaining errors are our own.
1 Introduction

On August 23, 2016, BBC.com published a news story covering the ban of Russian athletes from the 2016 Paralympic games due to a doping scandal. The title of the story was “Rio Paralympics 2016: Russia banned after losing appeal,” and the news piece discussed that an arbitration court upheld an earlier decision to exclude the Russian team from the Paralympics.\(^1\) On the same day, another online news agency, RT.com, published a story about the same news event, titled “Removing a strong rival? Russia shocked by “cynical and political” CAS ruling on Paralympic team ban.”\(^2\) The article emphasized that the decision of the arbitration court was political and that there was no hard evidence of Russian athletes doping.

Such a difference in media coverage is defined by the economics literature as media slant (e.g. Gentzkow and Shapiro 2010). Media slant can come from the supply side, reflecting preferences of journalists, advertisers, owners of the news outlet, or governments (Baron, 2006; Besley and Prat, 2006). It can also come from the demand side, reflecting preferences of readers for like-minded news or diverse information sources (Mullainathan and Shleifer, 2005; Xiang and Sarvary, 2007; Gentzkow and Shapiro, 2010; Zhu and Dukes, 2015). In the case of the coverage of the Russian Paralympic team ban, there are reasons to suspect that media slant comes from the supply side, given that RT.com is owned by the Russian government, and “RT” is short for “Russia Today.” However, this government-induced bias would be different from the government capture in the traditional models (Besley and Prat, 2006), in which governments need to capture all news sources to suspend information dissemination. In the presence of competitors, does RT attract primarily the consumers whose ideological views are aligned with its reporting, in which case RT’s media slant might be demand-driven? Or does RT attract (and potentially persuade) customers whose ideological views are not aligned with its reporting, for example, by RT’s high quality or strong brand? If so, what fraction of RT’s readers have such misaligned views, and how does the government-induced slant affect RT’s market share?

In this paper, we aim to shed light on the above questions using the Russian online news market in the years 2013-2015 as a case study. In this period, the Russian online news market had both government-controlled (GC) and independent news outlets, with all news outlets being easily accessible on the internet. A stylized fact in this market is that the GC news outlets enjoyed high and stable market shares in this period of time: around 25% of

\(^1\)http://www.bbc.com/sport/disability-sport/37165427  
\(^2\)https://www.rt.com/sport/356863-paralympic-russia-reaction-rio/
the top news outlets were GC (owned by the government), and their overall market share was around 35-40\%\textsuperscript{3}. The key question of this paper is what drives demand for the GC news outlets in Russia. To formalize the motivating example of RT above, we distinguish between two potential families of explanations. First, consumers might have a preference for the ideological bias in the GC news coverage, either because their beliefs align with the government’s ideological position, they value knowing the ideological position of the government, or they find ideologically-heavy news content entertaining. What fraction of consumers prefer the ideological slant of the GC news outlets over the independent news outlets, and how does this ideological slant affect the market shares of the GC news outlets? Second, consumers might have a persistent preference for the GC news outlets, reflecting a potentially higher quality or brand capital of these outlets. Do consumers prefer the quality of the GC news outlets over the independent news outlets, and if so, by how much does the higher quality increase the attention share and media power of the GC news outlets?

To answer these questions, we collect two novel datasets. First, we collect all the publication records for the top 48 online news outlets that wrote in Russian during the period of April 2013 - April 2015. The resulting panel contains 3.9 million news articles, and for each article we know its URL, text, title and publication date. Second, we collect the browsing records for 284,574 consumers of Russian online news websites in the period of November 2013 - April 2015 from the Internet Explorer Toolbar dataset. This data provides us with a long panel of the instances of news consumptions.

We use the publication records data to find and characterize pro-government bias in the news. For this, we compare publications of the GC and independent news outlets, identities of which we know a priori from the ownership structure. To make this comparison, we design a novel and simple classification method that allows us to find the words and phrases that are under- or overused by most of the GC news outlets. To identify sensitive news, we use two potential methods of government control: censorship and propaganda. First, we use the idea of censorship and find topics (identified as proper nouns) that are systematically underused by the GC news outlets compared to the independent news outlets. There is a significant difference in coverage, indicating that the GC news outlets systematically omit news about political opposition and corruption, such as news about opposition leaders (e.g., “Khodorkovsky,” “Navalny”), affiliates of President Putin’s who are tied to corruption (e.g., “Rotenberg,” “Timchenko”), and political protests (e.g. “Bolotnaya”). We label censored

\textsuperscript{3}Based on the statistics from \url{http://www.liveinternet.ru}. Historical data is scrapped using the Wayback Machine: \url{http://web.archive.org/}.  

3
news as “internally sensitive news” and characterize the ideological positions of the news outlets as a share of internally sensitive news that they report. Second, we examine news coverage of the Ukraine crisis of 2013-2015, a major sensitive news topic for the Russian government. We show that the GC news outlets systematically report more news about the Ukraine crisis compared to the independent news outlets, but their news coverage exhibit a pro-Russia (anti-Ukraine) slant; for example, GC news outlets tend to say that Crimea has “reunited” with Russia, the new Ukrainian government is fascist and “anti-Russian,” and the Ukrainian government is conducting a “punitive” operation against rebels in eastern Ukraine. Similarly, the Ukrainian news outlets with coverage in Russian systematically use a pro-Ukraine (anti-Russia) slant; for example, they report that Russia has “annexed” Crimea, Russia is an “aggressor” country, and the Ukrainian government is conducting an “anti-terrorist” operation against “terrorists” in eastern Ukraine. Using these vocabularies of ideologically-slanted words, we construct measures of valence and volume of the ideological slant and characterize the ideological positions of the news outlets.

Having the ideological positions of the news outlets, we build and estimate a demand model for news. Consumers have persistent preferences for news outlets, preferences over the news topics and preferences for the ideological slant in sensitive news coverage. Persistent preferences capture the attitude of consumers towards the fixed characteristics of the news outlet, such as the overall quality of the outlets and their brand, which we will refer to as “quality” from now on. To separate out the persistent preferences of consumers from their preference for news topics and ideological slant, we use the variation in the amount of sensitive news over time, which we measure with the overall daily share of publications about the sensitive news. On days with no sensitive news, ideological differences between the GC and independent news outlets do not affect consumers’ utilities of reading the news, which is driven only by the persistent preferences for news outlets. In contrast, on days with a lot of sensitive news, there are ideological differences between the GC and independent news outlets, and consumer preferences for ideological content become important. If consumers switch to reading the GC news outlets on days with sensitive news, they prefer a pro-government bias in the news. If they switch away from the GC news outlets on these days, they have a distaste for pro-government bias. If consumers start reading more ideologically-diverse news outlets on the days with a lot of sensitive news, they behave like “conscientious” news readers who value knowing the alternative sides of the story (Mullainathan and Shleifer, 2005), whereas more ideologically-similar news diet would suggest more of confirmation-bias-driven news readers. Finally, consumers might care only about the volume of the slant, for
example, if they consume the slanted news only for the entertainment purposes, in which case they would be more likely to visit any slanted news outlet, regardless of the valence of its slant.

Estimates of the structural model reveal a nuanced picture. On the one hand, an average consumer in our sample prefers the ideological position of the independent news outlets and the quality of the GC news outlets, suggesting that quality is the primary driver of the demand for the GC news outlets. On the other hand, there is substantial heterogeneity in consumer preferences, with a segment of consumers having a preference for the ideological position of the GC news outlets. In particular, 37.2% of consumers prefer the ideological slant of the GC news outlets in coverage of the Ukraine crisis, and 42.11% of consumers prefer the coverage of the GC news outlets about the internally sensitive news.

To test whether the GC news outlets lose market share because of their ideological positions, we perform several counterfactual simulations, changing the ideological positions of the news outlets and their quality level. If the ideological position of the GC news outlets was similar to the independent news outlets, the GC news outlets would get a 20.3% higher market share, corresponding to a rough back-of-the-envelope estimate of $18.41 million in advertising revenues. This loss pales in comparison to $1.21 billion of government subsidies to mass media in Russia in 2015 alone, suggesting that it is relatively easy for the government to cover the controlled outlets’ advertising losses. In contrast, if the average quality of the GC news outlets was similar to the independent news outlets, they would get a 44.6% lower market share, showing that quality is the primary reason for the observed demand for the GC news outlets.

In light of the importance of quality in the demand for the GC news outlets in this market, we discuss the implied media power (Prat, 2017), ability of a set of the news outlets to swing the elections, induced by the high quality of the GC news outlets. We show that the higher quality of the GC news outlets increases their share of online attention from 21.78% to 33.4%, substantially increasing their media power. On the days with a lot of sensitive news, such “brand media power” allows the GC news outlets to capture 21.5%-31.1% of the online attention of consumers who would prefer the coverage of the independent news outlets.

Finally, structural demand estimates allow us to separate out alternative mechanisms behind the demand for biased news coverage. The vast majority of consumers, 74.5%, prefer more ideologically-similar news sources on days with more Ukraine-crisis news. Thus, only

---

4Source: [http://www.rbc.ru/politics/29/06/2015/55912ff9a79474533f82cda9](http://www.rbc.ru/politics/29/06/2015/55912ff9a79474533f82cda9). We use the exchange rate of the end of 2014, which was 60 rubles per dollar. The total of 72.6 billions rubles includes subsidies to the television and print media.
a minority (25.5%) of consumers behave like “conscientious” news readers, and an average consumer prefers news with less volume of slant, suggesting that preference for like-minded news is the main driver behind the demand for ideologically-slanted news outlets.

This paper is the first to use a structural demand model to estimate consumer preferences for pro- and anti-government slant in autocracies.\(^5\) We propose a new and simple method to measure media slant (Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010; Gentzkow et al., 2016), and to our knowledge we are the first to separate out media bias into censorship, valence and volume of slant.\(^6\) We use a new identification strategy to estimate consumer preferences for slant in online news, which contributes to the empirical literature on media slant (Gentzkow and Shapiro, 2010; Martin and Yurukoglu, 2017) and online news markets (Gentzkow et al., 2011; Gentzkow and Shapiro, 2015; Sen and Yildirim, 2016; Athey et al., 2017; Cagé et al., 2017). Our demand estimation results inform theoretical literature on the demand-side slant (Mullainathan and Shleifer, 2005; Xiang and Sarvary, 2007; Zhu and Dukes, 2015). Our work is also related to the theoretical literature on government capture (Besley and Prat, 2006; Petrova, 2008; Prat and Strömberg, 2013; Edmond, 2013; Gehlbach and Sonin, 2014), empirical literature on the effect of government control of the news on consumers (Durante and Knight, 2012; Enikolopov et al., 2011; Bai et al., 2015; Roberts, 2014; Garcia-Arenas, 2016; Knight and Tribin, 2016), and to the literature on media power (Prat, 2017; Kennedy and Prat, 2017).

The rest of the paper is organized as follows. Section 2 describes the Russian online news market and data sources. In Section 3, we use the publication records data to find government-sensitive news topics and characterize the reporting of news outlets. Section 4 uses the variation in the amount of sensitive news over time to show the descriptive evidence. We build a demand model in Section 5. Section 6 discusses the estimation procedure and the results. We run the counterfactual simulations of changing the level of government control in Section 7. Section 8 concludes.

\(^5\)Gentzkow and Shapiro (2015) discuss a related demand model for online news consumption. Other work examined the consumer response to an increase in the pro-government bias in the news, with Durante and Knight (2012) documenting viewers’ response to the ideological change in TV programming of public television due to Berlusconi’s victory in the national elections in Italy and Knight and Tribin (2016) documenting viewers’ response to the airing of cadenas, government propaganda on Venezuela channels.

\(^6\)Perego and Yuksel (2016) discuss the separate decision of news outlets on agenda setting and slant in the news in a theoretical framework, and Pan and Xu (2017) examine if the Chinese ideological spectrum is multi-dimensional.
2 Data

2.1 Online News Market in Russia

We start with a brief overview of the Russian online news market for the period of our study. Despite relatively high government control over the offline news market starting in 2000, online news outlets in Russia enjoyed relative freedom up until 2013. A large number of independent players existed in the online news media landscape. Since the beginning of 2013, political pressure has forced a number of top online news outlets to remove their chief editors.\(^7\) The most prominent examples include the dissolution of RIA Novosti, a state news agency known for balanced news coverage under its editor-in-chief Svetlana Mironyuk, in December 2013\(^8\) and the layoff of Galina Timchenko, editor-in-chief of one of the top online news outlets in Russia, lenta.ru, in March 2014.\(^9\) Government control intensified in February of 2014 with the Ukrainian crisis and the annexation of Crimea, with the government blocking websites of some opposition leaders in March 2014\(^10\) and implementing a law to limit the foreign ownership of Russian news outlets to 20% as of January 2016.\(^11\)

At the end of 2014, the online news media landscape in Russia still included both groups of GC and independent news outlets. Table 1 presents the top 48 Russian-language news outlets,\(^12\) including 40 Russian news outlets, five international news outlets that offer news stories in Russian, and three large Ukrainian outlets with Russian-language versions of the websites. Russian news outlets are organized into four groups: (1) independent and uninfluenced, (2) independent but possibly influenced, (3) possibly influenced news media owned by oligarchs close to the Kremlin, and (4) GC outlets. Classification is based on the interviews with media professionals who prefer to remain anonymous. The ownership structure and media reports support this classification. For example, news outlets classified as GC are owned

---


\(^12\)Top outlets are defined using their market share on liveinternet.ru.
Table 1: Russian-language online news media by the type of influence in December 2014

<table>
<thead>
<tr>
<th>International</th>
<th>Independent</th>
<th>Possibly Influenced (oligarchic)</th>
<th>Government (controlled)</th>
<th>Ukrainian (subset)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bbc</td>
<td>newsru</td>
<td>bfm</td>
<td>fontanka</td>
<td>1tv</td>
</tr>
<tr>
<td>svoboda</td>
<td>newtimes</td>
<td>echo</td>
<td>gazeta</td>
<td>aif</td>
</tr>
<tr>
<td>meduza</td>
<td>novayagazeta</td>
<td>interfax</td>
<td>lifenews</td>
<td>dni</td>
</tr>
<tr>
<td>dw</td>
<td>rbc</td>
<td>mk</td>
<td>izvestia</td>
<td>ntv</td>
</tr>
<tr>
<td>reuters</td>
<td>slon</td>
<td>znak</td>
<td>kommersant</td>
<td>rg</td>
</tr>
<tr>
<td>tvrain</td>
<td>ng</td>
<td>kp</td>
<td></td>
<td>ria</td>
</tr>
<tr>
<td>vedomosti</td>
<td>polit</td>
<td>lenta</td>
<td></td>
<td>rt</td>
</tr>
<tr>
<td>forbes</td>
<td>regnum</td>
<td></td>
<td></td>
<td>vesti</td>
</tr>
<tr>
<td>snob</td>
<td>ridus</td>
<td></td>
<td></td>
<td>vz</td>
</tr>
<tr>
<td>the-village</td>
<td>rosbalt</td>
<td></td>
<td></td>
<td>tass</td>
</tr>
<tr>
<td></td>
<td>sobesednik</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>utro</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>trud</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table presents the simplified domain names; for example, 1tv stands for www.1tv.ru. Most domains have the www.*.ru structure, with some exceptions. Groups are created based on the publicly available information about ownership structure and interviews with media professionals.

by the government (7 out of 10 news outlets) or the state company Gazprom (1 news outlet), or were founded by a member of the current incumbent party and a strong supporter of Vladimir Putin, Konstantin Rykov (2 news outlets). Appendix 10.1 contains detailed information on the ownership structure and public information about the news outlets.

Functionally, group (1) of independent and not-influenced news outlets does not experience direct government influence, but might be subject to self-censorship given that the Russian government can potentially punish the news outlets. Groups (2) and (3) of independent but potentially influenced news media and oligarchic media have officially independent news coverage, but are reported to be indirectly influenced by the government. The nature of government control in these groups is very similar, so we group these outlets together and call them “influenced” news outlets. Finally, group (4) contains GC news outlets that have a news agenda directly controlled by the Kremlin. The majority of these news outlets are owned by the government and receive government subsidies.

8
2.2 Supply Data

For the 48 outlets described above, we collect information on publications for the period starting April 1, 2013, and ending March 31, 2015. The data are collected directly from archives on news-outlet websites and from the media archives mediologia.ru and public.ru. The resulting panel contains 3.9 million publications. For each article, we collect the title, text, URL link, and timestamp. Table 2 presents the number of articles per type of news outlet. Appendix 10.2 provides more information on the publication records data collection and processing.

Table 2: Number of articles by type of news outlet

<table>
<thead>
<tr>
<th>Type</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC</td>
<td>1,168,569</td>
</tr>
<tr>
<td>Independent</td>
<td>494,087</td>
</tr>
<tr>
<td>Influenced</td>
<td>1,848,556</td>
</tr>
<tr>
<td>International</td>
<td>75,596</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>315,927</td>
</tr>
<tr>
<td>Total</td>
<td>3,902,735</td>
</tr>
</tbody>
</table>

2.3 Demand Data

To measure news consumption, we use the Internet Explorer (IE) Toolbar browsing data, which includes complete browsing histories for a subset of IE users. The users included in the IE Toolbar data have installed a plug-in on their IE and opted-in for the data collection.\textsuperscript{14} IE Toolbar data contain information about each webpage consumers visited (URL), websites where consumers came from (referral URL), timestamp of the visit, number of seconds spent, browsing session ID, user ID, language of the browser, country of the user, and other information. We focus the analysis on Toolbar users who specified Russian as the language of their browser.\textsuperscript{15}

Although IE Toolbar data are collected for several years, the unique user IDs are kept only for one and a half years. By the time the data collection was conducted, the earliest

\textsuperscript{13}For five news outlets (“meduza,” “newtimes,” “ridus,” “snob,” “the-village”), text was not collected for technical reasons. We keep these outlets in parts of the textual analysis and use titles instead. We drop these news outlets for the descriptive analysis and demand estimation because without information on article texts, we could not get a reliable measure of slant.

\textsuperscript{14}Based on Microsoft records, around 75% of users who installed the plug-in opt-in to the data collection.

\textsuperscript{15}Having a browser in Russian language indicates that the user knows Russian and is potentially in the market for Russian online news.
available browsing data with user IDs were from November 15, 2013. We thus collect the browsing data between November 15, 2013, and March 31, 2015\textsuperscript{16} for all users with the IE language set to Russian.

The resulting panel consists of 2.17 million users. Among these users, 284,574 navigated to a news website at least once over the sample period, which is only 13\% of users with IE in Russian. At the same time, these users are the most active online; their browsing corresponds to 77.8\% of all browsing of users who set their IE language to Russian. In total, our sample contains 26.54 million page views of the 48 news-outlet websites defined above.

To understand the online news consumption in the IE Toolbar data, we classify the webpages that consumers visit into four groups: main pages of the websites, news subdirectories, news articles, and other pages (such as special projects, photos, videos, etc.).\textsuperscript{17} Table 13 shows summary statistics of browsing by types of webpages. News articles account for 39.3\% of the page views on news websites. News directories and subdirectories account for another 36\%, with other pages accounting for 24.7\%.

<table>
<thead>
<tr>
<th></th>
<th>Page views</th>
<th>% of Page Views</th>
<th>Seconds spent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Main page</td>
<td>5,344,041</td>
<td>20.1%</td>
<td>128</td>
</tr>
<tr>
<td>News articles</td>
<td>10,420,780</td>
<td>39.3%</td>
<td>186</td>
</tr>
<tr>
<td>News subdirectories</td>
<td>4,225,221</td>
<td>15.9%</td>
<td>263</td>
</tr>
<tr>
<td>Other</td>
<td>6,584,713</td>
<td>24.7%</td>
<td>145</td>
</tr>
<tr>
<td>Total</td>
<td>26,537,267</td>
<td>100%</td>
<td>176</td>
</tr>
</tbody>
</table>

### 2.3.1 IE Toolbar Representativeness

Before we proceed with the analysis, we test whether the news consumers in the IE Toolbar data are representative of the overall population of news consumers in Russia. To make this comparison, we collect data on the number of daily visits for a subset of news outlets in Russia using liveinternet.ru (LI), a website that tracks statistics for the Russian internet. We use the digital archive Wayback Machine to collect historical data on website usage. Due to the layout of the website ranking on LI, we have reliable records of usage over the period

\textsuperscript{16}Data for the period between April 1, 2013, and November 15, 2013, are available with scrubbed (deleted) user IDs.

\textsuperscript{17}Appendix 10.3 contains details on classification.
of time studied for the top 30 websites on the Russian internet, which includes seven news websites from our sample.\textsuperscript{18}

First, in Table 4 we compare the visit shares and rankings of the news outlets in the IE Toolbar and LI data. Results are mixed. On the one hand, five out of the top seven news outlets in the LI data are also present in the top seven in the IE Toolbar data. On the other hand, there are a couple of significant deviations, with the second outlet in the LI data, ria.ru, ranking 14 in the IE Toolbar data, and the market leader, rbc.ru, having a substantially higher visit share in the IE Toolbar data. One of the potential explanations for these differences is the anecdotal over-representativeness of the office workers in the IE Toolbar data. This can explain both a higher visit share of the rbc.ru (it is a news agency with a more extensive business news coverage) and a lower visit share of the ria.ru (it is another news agency competing with rbc.ru). Table 14 in the Appendix 10.4 compares the other browsing habits of the IE Toolbar data users and LI users and indeed shows some suggestive evidence that the IE Toolbar data over-represent the office workers; users of IE Toolbar are less likely to navigate to the entertainment websites and are likely to be older.\textsuperscript{19}

<table>
<thead>
<tr>
<th>Website</th>
<th>liveinternet.ru Visit Share</th>
<th>IE Toolbar Visit Share</th>
<th>Rank</th>
<th>IE Toolbar Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>rbc.ru</td>
<td>0.1951</td>
<td>0.3165</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ria.ru</td>
<td>0.1800</td>
<td>0.0570</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>vesti.ru</td>
<td>0.1550</td>
<td>0.1879</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>kp.ru</td>
<td>0.1355</td>
<td>0.1146</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>lenta.ru</td>
<td>0.1319</td>
<td>0.1094</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>gazeta.ru</td>
<td>0.1248</td>
<td>0.1010</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>rg.ru</td>
<td>0.1240</td>
<td>0.1135</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

IE Toolbar rankings are computed out of the 48 news outlets described in Table 1.

Second, we examine changes of the news outlets’ traffic in the IE Toolbar and LI datasets. Figure 1 presents the average traffic to the top seven LI ranking news outlets based on the LI and IE Toolbar data.\textsuperscript{20} Changes in the news consumption in the IE Toolbar data closely track the population-level consumption in the LI data, with the correlation of 0.858. Figure 13

\textsuperscript{18}The top page includes only the top 30 websites; Wayback Machine does not have frequent records for the other pages.

\textsuperscript{19}See the Appendix 10.4 for a more detailed comparison.

\textsuperscript{20}For each website and news source, the average traffic level is normalized to one, and the IE Toolbar data are corrected for the churn rate. The traffic is then averaged across the news outlets. Figure 14 in the Appendix 10.5 contains information about the number of weekly users of the IE Toolbar.
in the Appendix 10.4 presents changes in the traffic for each of the top seven news outlets. The correlation between traffic changes in the LI and IE Toolbar datasets vary from 0.52 to 0.914. Overall, changes in the consumption of news in the IE Toolbar data track the changes in the consumption of the population well, even for the websites with substantial differences in the average market shares, rbc.ru (correlation of 0.914) and ria.ru (correlation of 0.702).

Figure 1: Normalized number of weekly visitors to rbc.ru, IE Toolbar and LI data

For each website and news source, the average traffic level is normalized to one, and the IE Toolbar data are corrected for the churn rate. The traffic is then averaged across the news outlets.

3 Government Control and Sensitive News

3.1 Types of Government Control

In this section we characterize the government control and pro-government bias in the online news market in Russia.

In general, researchers acknowledge two broad types of news bias induced by governments: censorship and propaganda. Censorship of the news occurs when the government removes a certain topic of fact from the news coverage of its controlled outlets. For example, a government instructing a news outlet not to cover a story about a corruption scheme organized by some government officials or instructing a news outlet to omit certain facts about the involvement of government officials in the scheme is classified as censorship. The media economics literature refers to censorship as “issue and fact bias” (Prat and Strömberg, 2013) or as “filtering or selection of news” (Gentzkow et al., 2016). Censorship works through
the effects of agenda setting (McCombs and Shaw, 1972) and priming (Iyengar and Kinder, 1987). Cohen (1963) summarized the idea of agenda setting by arguing that the press “may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to think about.”

Apart from censorship, government can control the news by adding slant to news reporting. We will refer to such slant as propaganda and will use the words “slant” and “propaganda” in this work interchangeably. We define propaganda as news reporting with language favoring one of the parties described in the news. Gentzkow and Shapiro (2010) provide multiple examples of slanted language used by the members of the US Congress, such as describing the Iraq war as “war on terror” (republicans) versus “war in Iraq” (democrats). In the media economics literature, propaganda corresponds to “framing and ideological stand bias” (Prat and Strömberg, 2013) and “distortion of news” (Gentzkow et al., 2016).

Throughout this paper, we treat the news product that outlets offer as a combination of three components: news topics that are reported, slant in sensitive news topics, and a news outlet’s quality that represents the persistent characteristics of the outlet. Censorship affects the topics that the GC news outlets report, and government propaganda affects the degree of pro-government slant in the sensitive news topics reporting.

### 3.2 Internally Sensitive News

We use the definitions of government control and publication records data to find sensitive news topics. To do this, we treat news articles as collections of words, simplified with standard natural language processing techniques such as stemming and removal of stop words.

We identify the first set of government-sensitive news using the idea of censorship. Given that censorship is the omission of facts, we focus on proper nouns, words that are likely to correspond to facts in the news. For example, proper nouns represent the actors in the news and the places where the news happened. We consider all words starting with

---

21 We note that our definition of slant is more narrow than the definition of Gentzkow and Shapiro (2010) as we define censorship and slant separately.

22 Appendix 10.2 provides more information on processing of the publication records data.

23 Franceschelli (2011) also uses proper nouns to define the news topics. Cagé et al. (2017) uses a more sophisticated topic detection algorithm.

24 For example, a title of one of top news stories on the day when this paragraph was written, “Panama Paper: David Cameron’s worst week as Prime Minister,” contains proper nouns “Panama Papers,” “David Cameron,” and “Prime Minister,” which summarize the topic of the news article, but do not capture the sentiment of this topic (captured by the word “worst”).
a capital letter as proper nouns except for the first words in the sentences. If the facts corresponding to a topic are censored or underreported, proper nouns related to the topic will be underused in the news outlet’s publications.

Censored proper nouns should be underused by all the GC news outlets compared to all the independent news outlets. To find such proper nouns, we design a simple detection algorithm:

1. Compute share of usage of a word $v$ by a news outlet $j$: $\text{sh}_{vj} = \frac{\text{count}_{vj}}{\sum_v \text{count}_{vj}}$ for all $v, j$,

2. For each $v$, rank $\text{sh}_{vj}$ across the news outlets $j \in \{1, \ldots, 48\}$:
   - $\text{rank}_{vj}’ = 1$ if $\text{sh}_{vj}’ = \max_j (\text{sh}_{vj})$
   - $\text{rank}_{vj}'' = 2$ if $\text{sh}_{vj}'' = \max_j (\text{sh}_{vj}) : j \neq j' (\text{sh}_{vj})$
   etc.,

3. For each $v$, compute an average rank for the GC and independent news outlets:
   - $\text{Rank}_v^x = \frac{\sum_{j \in x} \text{rank}_{vj}}{\sum_{j \in x}}$.

The proper nouns that are censored by the GC news outlets should have low $\text{Rank}_v^{\text{Gov}}$ and high $\text{Rank}_v^{\text{Ind}}$, so proper nouns with negative rank difference $\Delta \text{Rank}_v^{\text{Ind} - \text{Gov}} = \text{Rank}_v^{\text{Ind}} - \text{Rank}_v^{\text{Gov}}$ are likely candidates for sensitive censored news topics. To test if low values of $\Delta \text{Rank}_v^{\text{Ind} - \text{Gov}}$ occur by chance, we randomly re-assign word counts within the news outlets and re-do the procedure. Distribution of $\Delta \text{Rank}_v^{\text{Ind} - \text{Gov}}$ from the re-assigned scenario provides a benchmark distribution of the differences in ranks. We define the empirical distribution of $\Delta \text{Rank}_v^{\text{Ind} - \text{Gov}}$ based on actual corpus as $h_{\text{ind-gov}}^{\text{actual}}$ and the empirical distribution based on re-assigned corpus as $h_{\text{ind-gov}}^{\text{random}}$.

We apply the above procedures to the 21,709 unigrams and 13,514 bigrams of the proper

---

25 We note that such measure of censorship does not account for self-censorship of the news outlets and corresponds to state censorship (Crabtree et al., 2015).

26 For example, if a news outlet used only three words A, B, and C, and these words were used count$_A = 10$, count$_B = 15$ and count$_C = 20$ times, random re-assignment of word counts within a news outlet will permute the observed counts, for example count'$_A = 20$, count'$_B = 10$, count'$_C = 15$. In the data, news outlets use tens of thousands of unique words, so an empirical distribution of the word counts should be a good approximation of an actual distribution of the words counts for a given outlet.

27 The benefit of this procedure as opposed to a simple comparison of shares of usage is twofold. First, news outlets differ in news volume; using shares of words instead of counts allows normalizing the size of news outlets. Second, some outlets might specialize in a particular topic (e.g., corruption scandals) and have a large share of usage of particular proper nouns. Using an ordinal-rank measure instead of cardinal-share measure for usage of proper nouns allows us to limit the effect of such outliers.
nouns that appear more than 200 times in the news publications in our sample period.\textsuperscript{28} Figure 2 presents the histograms of the rank difference distributions for bigrams, $h_{\text{actual}}^{\text{ind-gov}}$ and $h_{\text{random}}^{\text{ind-gov}}$.\textsuperscript{29} Bigrams on the left side of the $h_{\text{actual}}^{\text{ind-gov}}$ distribution (negative rank difference) correspond to the potentially censored proper nouns. To test if these differences in ranks could occur by chance, we compare $h_{\text{actual}}^{\text{ind-gov}}$ to $h_{\text{random}}^{\text{ind-gov}}$. The lowest rank difference in $h_{\text{actual}}^{\text{ind-gov}}$ is -28.3, while in $h_{\text{random}}^{\text{ind-gov}}$ it is -18.8, corresponding to the red dashed line in the Figure 2.\textsuperscript{30} Thus, the bigrams in the actual corpus with rank difference of less than -18.8 are systematically underused by the GC news outlets, and this difference cannot be explained by chance.

Figure 2: Histograms of $\Delta \text{Rank}_{\text{Ind-Gov}}$ across bigrams of the proper nouns: actual and random corpus.

The histogram in blue color corresponds to the actual corpus; the histogram in green color – to the random corpus. The red vertical line is a cutoff corresponding to the lowest rank difference in the random sample, -18.8.

We use the lowest rank difference in $h_{\text{random}}^{\text{ind-gov}}$, -18.8, as the threshold to define censored proper nouns. There are 54 bigrams of the proper nouns in the actual sample with the rank difference below this threshold. After excluding proper nouns related to the profession of

\textsuperscript{28}The threshold of 200 times ensures that the proper nouns in the analysis refer to the substantial topics and is chosen arbitrarily.

\textsuperscript{29}Appendix 10.6 presents the unigram results.

\textsuperscript{30}To make sure that the difference in the lowest $\Delta \text{Rank}_{\text{Ind-Gov}}$ in the actual and random corpora did not occur by chance, we repeat the re-assignment procedure 500 times and use the 5% quantile of the lowest rank difference across the re-assignments as the lowest rank difference in $h_{\text{random}}^{\text{ind-gov}}$. 

15
journalism (such as journalists, news outlets, media owners, etc.), we are left with 34 bigrams of the proper nouns that we define as censored. Table 5 provides an example of these proper nouns by presenting 10 bigrams with the lowest $\Delta$Rank$_{Ind-Gov}$ difference. All of the bigrams are actors related to facts that are anecdotally known to be sensitive for the government: there are three prominent opposition politicians, two close affiliates of Vladimir Putin who are frequently mentioned in the events of potential corruption, Pussy Riot, a band that became famous for its protest activities, Sergei Guriev, a prominent Russian economist who had to flee Russia after a politically-motivated interrogation, Svetlana Davydova, a mother-of-seven arrested for a phone call to the Ukrainian embassy that was claimed to be an act of treason, and Marat Gelman, a former director of the Perm Museum of Contemporary Art allegedly fired for refusing to remove a controversial political exposition “Welcome! Sochi 2014”.

Table 5: List of the top 10 bigrams of the proper nouns underused by GC news outlets.

<table>
<thead>
<tr>
<th>Underused proper noun: English translation</th>
<th>Information about the proper nouns</th>
<th>Rank Difference, $\Delta$Rank$_{Ind-Gov}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexei Navalny</td>
<td>Opposition politician</td>
<td>-28.3</td>
</tr>
<tr>
<td>Mikhail Khodorkovsky</td>
<td>Opposition politician, political prisoner</td>
<td>-26.7</td>
</tr>
<tr>
<td>Sergei Guriev</td>
<td>Economist, interrogated about “Yukos”</td>
<td>-25.8</td>
</tr>
<tr>
<td>Gennady Timchenko</td>
<td>Businessman, close friend of Vladimir Putin</td>
<td>-25.7</td>
</tr>
<tr>
<td>Svetlana Davydova</td>
<td>Civilian, investigated for treason</td>
<td>-24.6</td>
</tr>
<tr>
<td>Marat Gelman</td>
<td>Gallerist, fired for a political exposition</td>
<td>-24.4</td>
</tr>
<tr>
<td>Alexei Navalny (2)</td>
<td>Opposition politician</td>
<td>-24.3</td>
</tr>
<tr>
<td>Ilya Yashin</td>
<td>Opposition politician</td>
<td>-24</td>
</tr>
<tr>
<td>Pussy Riot</td>
<td>Protest punk rock band</td>
<td>-23.2</td>
</tr>
<tr>
<td>Arkady Rotenberg</td>
<td>Businessman, close friend of Vladimir Putin</td>
<td>-22.3</td>
</tr>
</tbody>
</table>

We use the underused bigrams of the proper nouns to define the first group of sensitive news. To make sure that we do not exclude facts described with a single proper noun, we re-do the classification procedure using unigrams of proper nouns and add an additional 10 censored proper nouns based on that classification. We call the news related to these censored bi- and unigrams of proper nouns “internally sensitive” because most of the censored proper nouns correspond to the internal issues such as political opposition, protests and corruption. We define an article that contains at least one of the censored bi- and unigrams

---

31 Tables 15 and 16 in the Appendix 10.6 present all 54 censored bigrams.
32 Alexei Navalny appears there twice spelled in two different ways.
34 See the Appendix 10.6 for more details.
as an article about an internally sensitive news topic.\footnote{We note that our measure of censorship does not include potential self-censorship as it is based on the observed difference in language between the GC and independent outlets.}

### 3.2.1 Slant in the Internally Sensitive News

We have found the internally sensitive topics using the mechanism of censorship. However, we do not know if it is the only mechanism that the government uses to bias these sensitive news stories. On the occasions when internally sensitive news are not censored, the GC news outlets might report them with a slant different from the independent news outlet.

To test if such a slant exists, we examine whether the GC news outlets use different language in the articles about the internally sensitive news compared to the independent news outlets. We construct a corpus based only on the articles about the internally sensitive topics and apply a ranking procedure described above to words that are not proper nouns.

Figure 3 presents the histogram of $h_{\text{ind-gov}}^{\text{actual}}$ and $h_{\text{ind-gov}}^{\text{random}}$ for this corpus. The two distributions are almost identical. We use the cutoff mechanism described above to test if there is some systematic slant in the internally sensitive news and find little evidence for this; out of 37,734 words in the corpus, only four words are systematically omitted by the GC news outlets, and only one word out of them is indicative of slant (the word “prisoner” related to the arrested opposition activists).\footnote{Among the other three unigrams are the words “interview” and “editor,” related to the means of information delivery, and the word “fired,” related to the event with firing one of the journalists of an independent news outlet. We exclude these words as they are related to the journalism profession and not to the news covered.}

We conclude that we do not find evidence that the slant in the coverage of internally sensitive news by the independent and GC news outlets is different.

### 3.3 News about the Ukraine Crisis

We now turn to another government-sensitive news topic: the Ukraine crisis of 2013-2014 with a following conflict between Russia and Ukraine.\footnote{For a broad overview, see https://en.wikipedia.org/wiki/Ukrainian_crisis.}

The conflict was widely covered in the Russian news media and was reported to be heavily slanted by news outlets controlled by the Russian government.\footnote{For an overview, see https://en.wikipedia.org/wiki/Media_portrayal_of_the_Ukrainian_crisis#Media_in_Russia.}

We first examine if news about the Ukraine crisis were censored by the GC news outlets. To measure this, Figure 4 presents the share of news articles that contain the word “Ukraine”
Figure 3: Histograms of $\Delta \text{Rank}_{\text{Ind}-\text{Gov}}$ across words in the internally sensitive news topics corpus.

The histogram in blue color corresponds to the actual corpus; the histogram in green color – to the random corpus. The red vertical line is a cutoff corresponding to the lowest rank difference in the random sample, and the blue vertical line is a cutoff corresponding to the highest rank difference in the random sample.
that were published in the independent, government-influenced, and GC outlets over time.\textsuperscript{39} With the beginning of the Ukraine conflict, all types of the news outlets increased their reporting about Ukraine, but GC outlets increased it more than independent and influenced outlets.

Figure 4: Share of articles containing the word “Ukraine” in the weekly coverage of news outlets by types.

The red line corresponds to the GC media, the green line - to the independent media, and the blue line - to the government-influenced media. The red dotted line corresponds to February 22, 2014, the day when the former president Yanukovych fled Ukraine as a result of a revolution. The blue dotted line corresponds to the first Minsk Peace agreement, September 4, 2014.

We next look for the evidence of media slant in the Ukraine-crisis news. As before, we treat any article that contains the proper noun “Ukraine” as being related to the news about the Ukraine crisis. To find media slant, we compare the words used in the publications about the Ukraine crisis by the GC news outlets and the Ukrainian news outlets. This is motivated by the anecdotal evidence that news coverage in the Ukraine conflict suffers from both pro-Russia and pro-Ukraine media slant.\textsuperscript{40} Anecdotally, the pro-Russian slant frames the new Ukrainian government as a “fascist junta” that is conducting a “punitive operation” against the “rebels” in the Eastern Ukraine, and the pro-Ukraine slant frames Russia as an “aggressor” that has “occupied” the Ukrainian territory and supports “terrorists” and “separatists” in the Eastern Ukraine.

\textsuperscript{39}Having the word “Ukraine” in the news coverage is a proxy for an article being about the Ukraine crisis.

\textsuperscript{40}For example, the difference in the media slant is discussed on the fact-checking website stopfake.org, supported by faculty and alumni of the Mohyla School of Journalism and students from the Digital Future of Journalism program in Kyiv, Ukraine.
To test if there is a systematic difference in the reporting about the Ukraine crisis by the GC and Ukrainian news outlets, we construct a corpus based only on the articles about the Ukraine-crisis news topic and apply a ranking procedure described above to words that are not proper nouns. Figure 5 presents the histogram of $h_{\text{ukr-gov}}^{\text{actual}}$ and $h_{\text{ukr-gov}}^{\text{random}}$ for this corpus. We use the cutoff mechanism described above to test if there is some systematic slant in the Ukraine-crisis news coverage. Out of the 38,584 words in the corpus, there are 13 words that are systematically overused by the GC news outlets and only 2 words that are systematically under-reported by the GC news outlets compared to the Ukrainian news outlets. However, given that the $h_{\text{ukr-gov}}^{\text{actual}}$ and $h_{\text{ukr-gov}}^{\text{random}}$ distributions are different and that we know the anecdotal nature of the pro-Russia and pro-Ukraine slant, we can examine the overused words more broadly.

Figure 5: Histograms of $\Delta \text{Rank}_{\text{Ukr-Gov}}$ across words in the Ukraine-crisis news topic corpus.

![Histograms of $\Delta \text{Rank}_{\text{Ukr-Gov}}$ across words in the Ukraine-crisis news topic corpus.](image)

The histogram in blue color corresponds to the actual corpus; the histogram in green color – to the random corpus. The red vertical line is a cutoff corresponding to the lowest rank difference in the random sample, and the blue vertical line is a cutoff corresponding to the highest rank difference in the random sample.

Table 6 presents the top 10 overused words by the GC and Ukrainian news outlets in the Ukraine-crisis news coverage. Words overused by the GC news outlets are very consistent with the anecdotal evidence described above: they mention the “reunion” of Russia and Crimea, the “anti-Russian” “radical” protesters who have “overturned” the former government in a “coup,” and the “punitive” operation and “bombing” against the Eastern Ukraine
“rebels.” Words overused by the Ukrainian news outlets are more noisy but are still consistent with the anecdotal story from the above: Russia has “annexed” and “occupied” Crimea, and the Ukraine army is conducting an “anti-terrorist” operation against the “separatists.” We take this consistency as evidence that the overused words indeed correspond to a pro-Russia and pro-Ukraine slant. Using the overused words and the domain knowledge, we select 18 words that correspond to the pro-Russia slant and 7 words that correspond to the pro-Ukraine slant in the Ukraine crisis. Table 18 in the Appendix 10.7 contains the final list of the selected words. We denote the articles that contain both the word “Ukraine” and one of the selected pro-Russia- or pro-Ukraine-slanted words as an article about the Ukraine crisis with the pro-Russia or pro-Ukraine slant, respectively.

Table 6: List of the top 10 overused words by the GC and Ukrainian news outlets in the Ukraine-crisis news coverage.

<table>
<thead>
<tr>
<th>Overused words by the:</th>
<th>GC news outlets</th>
<th>Ukrainian news outlets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>$\Delta \text{Rank}_{\text{Ukr-Gov}}$</td>
<td>Word</td>
</tr>
<tr>
<td>reunion</td>
<td>34.7</td>
<td>continental</td>
</tr>
<tr>
<td>radical</td>
<td>34.1</td>
<td>annexation</td>
</tr>
<tr>
<td>punitive</td>
<td>33.5</td>
<td>monopolistic</td>
</tr>
<tr>
<td>overturn</td>
<td>33.1</td>
<td>anti-terrorist</td>
</tr>
<tr>
<td>blockade</td>
<td>32.6</td>
<td>devoid</td>
</tr>
<tr>
<td>bombing</td>
<td>32.2</td>
<td>titushky(^{43})</td>
</tr>
<tr>
<td>coup</td>
<td>31.7</td>
<td>content</td>
</tr>
<tr>
<td>anti-Russian</td>
<td>31.1</td>
<td>residue</td>
</tr>
<tr>
<td>colored</td>
<td>31</td>
<td>occupied</td>
</tr>
<tr>
<td>deepest</td>
<td>31</td>
<td>deduced</td>
</tr>
</tbody>
</table>

3.4 Reporting about the Sensitive News

3.4.1 Internally Sensitive News

Knowing the articles about internally sensitive news and slant, we now characterize the reporting of news outlets. Figure 6 presents the average share of articles about the internally sensitive topic by types of the news outlets. By construction, the independent news outlets report more about the internally sensitive news than the GC outlets. The influenced news

---

\(^{41}\) The word “rebels” is the fifty-eighth overused word with a rank difference of 27.3.

\(^{42}\) The word “separatists” is the twenty-first overused word with a rank difference of -28.4.

\(^{43}\) This is a Ukrainian word to describe protesters supporting the former Ukraine government.
outlets are in the middle, with some influenced news outlets being closer to the positions of the GC news outlet and some being closer to the independent news outlets. The position of the international news outlets is closer to the independent news outlets, and Ukrainian news outlets tend to report very little about the internally sensitive news, given that their coverage is focused on the issues in the Ukraine.44

Figure 6: Reporting about internally sensitive news, by news outlets’ types.

Each dot represents a position of a news outlet. We remove five news outlets for which we have only information about the titles and about the text of the articles.

In addition to the average positions of the news outlets, we explore how the reporting on sensitive news changes over time. In particular, we are interested in whether censorship is more prominent on the days with more internally sensitive news. To measure the volume of the internally sensitive news, we compute the share of articles about the internally sensitive news in the market for each day $t$ in the sample, $F_{t}^{IS}$. Figure 7 presents the changes in $F_{t}^{IS}$ over time. On average, 2.7% of articles are about the internally sensitive news. There is also a substantial variation in the volume of internally sensitive news across days, with $F_{t}^{IS}$ varying from 0.2% to 21.9%.

To access whether censorship is more prominent on the days with more internally sensitive news, we regress the difference in reporting about internally sensitive news between the GC and independent news outlets, $F_{Ind,t}^{IS} - F_{Gov,t}^{IS}$, on the share of articles about the internally sensitive news.

44 Appendix 10.8 presents the ideological positions for the labeled news outlets.
sensitive news in the market, $F_{t}^{IS}$. There is a significant positive correlation between $F_{Ind,t}^{IS} - F_{Gov,t}^{IS}$ and $F_{t}^{IS}$, indicating the GC news outlets indeed censor more articles on the days with a high volume of sensitive news. On average, the GC news outlets report around 30.3% of the amount of the internally sensitive news that the independent news outlets report, $E(\frac{F_{Gov,t}^{IS}}{F_{Ind,t}^{IS}}) = 0.303$. This ratio of reporting is constant in volume of the internally sensitive news.\(^ {45}\) We conclude that the average reporting of the news outlets about the internally sensitive news, $F_{j,t}^{IS}$, is a good proxy for their reporting about the internally sensitive news over time.

### 3.4.2 Ukraine-crisis News

We now examine the reporting of news outlets about the Ukraine-crisis news. Figure 8 presents the average share of articles about the Ukraine-crisis topic by types of the news outlets. Independent, influenced and GC news outlets report relatively the same about of news about the Ukraine crisis. International and Ukrainian news outlets report more about the Ukraine crisis.

Subfigure (a) in Figure 9 presents the average shares of articles about the Ukraine crisis that have at least one pro-Russia or pro-Ukraine slant word. By construction, the GC news outlets have a relatively high levels of pro-Russia slant and low levels of pro-Ukraine

\(^{45}\)See Subfigures (a) and (b) of Figure 19 in the Appendix 10.9.
Each dot represents a position of a news outlet. We remove five news outlets for which we have information only about the titles and about the text of the articles.

slant, and Ukrainian news outlets have the opposite ideological positions. International news outlets have a lot of pro-Ukraine slant and less of the pro-Russia slant. Independent news outlets have few articles that contain the pro-Russia slant and vary in the amount of articles with the pro-Ukraine slant. Influenced news outlets have few articles that contain the pro-Ukraine slant and vary in the amount of the pro-Russia slant.

Results in Subfigure (a) of Figure 9 show that the ideological positions of the news outlets cannot be captured by a unidimensional measure of the level of Russian propaganda. News outlets differ not only in the valence of slant that they demonstrate, which can be more or less pro-government biased, but also in the volume of slant, with some news outlets being more neutral than others. To capture these ideological positions, we define the valence of slant as the difference in the level of pro-Ukraine and pro-Russia slant and the volume of slant as the sum of pro-Russia and pro-Ukraine slant. Subfigure (b) presented the resulting volume \( (V_j^+) \) and valence \( (V_j^-) \) of the slant of the news outlets. The GC and some influenced news outlets have a negative valence of slant (corresponding to more pro-Russia slant in the reporting), while international, Ukrainian and some independent news outlet have a positive valence (more pro-Ukraine slant). The majority of independent and influenced news outlets

---

\(^{46}\)We normalize the mean and standard deviations of the measures of the pro-Russia and pro-Ukraine slant to make them comparable.
Figure 9: Ideological positions of the news outlets in the Ukraine-crisis news coverage.

Each dot represents a position of a news outlet. Subfigure (a) presents shares of pro-Russia- and pro-Ukraine-slanted articles about the Ukraine crisis. Subfigure (b) presents valence and volume of slant measured as a transformation of the measures of pro-Russia and pro-Ukraine slant.

is neutral in terms of the valence of slant and differ in its volume.

4 Descriptive Evidence

In the previous section, we have detected the government-sensitive news topics and have characterized the ideological positions of the news outlets’ reporting about these news topics. The results provide us with the necessary ingredients for building and estimating a demand model for news. However, before we get to the model, we present some descriptive evidence on the role of government control in consumer demand. In particular, we examine the relationship of news outlets’ market shares and the amount of sensitive news in the market. The ideological position of news outlets is more prominent on the days with more sensitive news. If consumers prefer the pro-government bias, on the days with more sensitive news the market shares of the pro-government biased news outlets should grow more compared to the market shares of the less pro-government biased news outlets.

We compute the market shares of the news outlets using the news consumption records in the IE Toolbar data. We define a news consumption of an outlet $j$ on day $t$ by consumer $i$ as navigation to at least one news article on this website by consumer $i$ on day $t$. If a
consumer is online on day $t$ but does not navigate to any news articles, we record that she has chosen an outside option of not consuming the news from one of the online outlets. To compute the market share of an outlet $j$ on day $t$, we sum up all news consumption of this outlet on day $t$ and divide it by a sum of total news consumption and outside option choices on this day.

We then examine changes in consumption due to an increase in the amount of sensitive news by regressing the market shares of news outlet $j$ on the amount of internally sensitive news events and Ukraine-crisis news events on day $t$:

$$\log(\text{share})_{jt} = b_{0j} + b^{IS}_j \log(F^{IS}_t) + b^{Ukr}_j \log(F^{Ukr}_t) + X'_{jt}d + \xi_{jt}$$

(1)

where $F^{IS}_t$ and $F^{Ukr}_t$ correspond to the share of articles about internally sensitive news and Ukraine-crisis news, and $X_{jt}$ corresponds to the controls, such as indicator variables for weekdays and time trends.\textsuperscript{47} The slope coefficients $b^{IS}_j$ and $b^{Ukr}_j$ correspond to the change in the market shares due to the change in the amount of sensitive news in the market.

We estimate $b^{IS}_j$ and $b^{Ukr}_j$ for 42 news outlets including weekday and week indicator variations as controls.\textsuperscript{48} Figure 10 summarizes and visualizes the estimation results. Each point on the subfigures (a)-(c) represents an estimate of $b^{IS}_j$ or $b^{Ukr}_j$ for the news outlet $j$. Points of larger size represent a larger absolute value of the estimates, with blue and red colors corresponding to positive and negative estimates of $b^{IS}_j$, respectively. Points with bold borders represent outlets with statistically significant estimates of $b^{IS}_j$.\textsuperscript{49}

Subfigure (a) of Figure 10 visualizes the estimates of $b^{IS}_j$. Results suggest that news outlets with more reporting about the internally sensitive news are more likely to get an increase in the market shares on the days with more news about the internally sensitive events. We test this more formally by regressing the $b^{IS}_j$ estimates on $\bar{F}^{IS}_j$, the average share of reporting about the internally sensitive events by the news outlets $j$. Table 7 presents the results of this regression based on $b^{IS}_j$ with the different level of controls in regression (1), $\hat{b}^{IS}_j = \hat{d}^{IS}_0 + F^{IS}_j \hat{d}^{IS}_1 + \hat{\xi}^{IS}_j$. In the specification with weekday and week fixed effect (column 4) that we’ve used above, the relationship between $\bar{F}^{IS}_j$ and $\hat{b}^{IS}_j$ is on the margin of being significant ($p < .05018$). In the three other specifications of regression (1) that are less restrictive (columns 1-3), the relationship between $\bar{F}^{IS}_j$ and $\hat{b}^{IS}_j$ is significant either on 5%.

\textsuperscript{47}In case of the observations with zero market share, we assign the lowest observed non-zero share of this outlet to this observation.

\textsuperscript{48}We exclude five news outlets for which we do not have information about the text of the articles, and 1 news outlets (\texttt{dw.de/ru}) for which we have few (10) news consumption occasions.

\textsuperscript{49}Significance is tested at 5% level; standard errors are heteroskedasticity and autocorrelation consistent.
Figure 10: Predicted changes in the news outlets’ market shares with the change in the amount of sensitive news by news outlet.

(a) Volume of internally sensitive news reporting

(b) Volume of Ukraine-crisis news reporting

(c) Slant in Ukraine-crisis news reporting

Each point represents a news outlet. The size of the points represents the degree of change of the market share of news outlets, measured as a percent of average market shares of this news outlet. The blue color corresponds to the increase in the market shares, and the red color corresponds to the decrease in the market share. The bold borders of the points correspond to significance of the change in the market share.
or is on the margin of significance. We interpret this as evidence that news outlets with more reporting about the internally sensitive news are more likely to get an increase in their market shares on the day with more sensitive news.

Table 7: Relationship between the estimates of the market share changes of news outlets, \( b_{IS}^j \), and their ideological positions on internally sensitive news \( \bar{F}_{IS}^j \).

<table>
<thead>
<tr>
<th>( F_{IS}^j )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{b}_{IS}^j )</td>
<td>0.107</td>
<td>0.124</td>
<td>0.252</td>
<td>0.301</td>
</tr>
<tr>
<td>( \hat{b}_{IS}^j )</td>
<td>(0.051)</td>
<td>(0.067)</td>
<td>(0.1)</td>
<td>(0.149)</td>
</tr>
</tbody>
</table>

Controls (from the regression 1):
- Weekday FE: N Y Y Y
- Time trend polynomial (4-order): N N Y N
- Week FE: N N N Y

Controls are included in the regression (1) estimating \( b_{IS}^j \). Standard errors are heteroskedasticity consistent.

Subfigures (b) and (c) of Figure 10 visualize the estimates of \( b_{Ukr}^j \). Results suggest that news outlets with more reporting about the Ukraine crisis news (subfigure b), lower pro-government valence of slant and higher volume of slant (subfigure c) are more likely to get an increase in the market shares on the days with more news about the Ukraine crisis. Similar to the case above, we test this relationships more formally by regressing the \( b_{Ukr}^j \) estimates on the average share of reporting about the Ukraine-crisis news, \( \bar{F}_{Ukr}^j \), valence of slant in the reporting, \( V_j^- \), and volume of slant, \( V_j^+ \). Table 8 presents the the results of this regression based on \( b_{Ukr}^j \) with the different level of controls in regression (1), \( \hat{b}_{Ukr}^j = d_0^{Ukr} + \bar{F}_{Ukr}^j d_1^{Ukr} + V_j^- d_2^{Ukr} + V_j^+ d_3^{Ukr} + \epsilon_{Ukr}^j \). Based on the specification with weekday and week fixed effect (column 4) that we’ve used above, there is statistically significant positive relationship between \( \hat{b}_{Ukr}^j \) and \( \bar{F}_{Ukr}^j \), \( V_j^- \) and \( V_j^+ \), supporting the claim that the news outlets that report more about Ukraine crisis and contain less pro-government propaganda and more slant overall are more likely to gain higher market shares during the days with a lot of news about the Ukraine crisis. However, the relationships between \( \hat{b}_{Ukr}^j \) and volume and valence of slant is more noisy in other specifications (columns 1-3).

We need to be careful with the interpretation of the results above. On the one hand, we can interpret the relationship between changes in the level of sensitive news over time and their market share as causal, under the conditional independence assumption (CIA) of the proxy for the level of sensitive news on day \( t \), \( \log(\text{share})_{jt} \perp \log(F^x_t)|X_{jt} \forall j, x = \{IS, Ukr\} \). This is a plausible assumption given that \( \log(F^x_t) \) is determined by sensitive news events.
happening on day \( t \), which is out of control of the market participants. Under the CIA, results in tables 7 and 8 indicate that an increase in the amount of sensitive news is more likely to lead to an increase in the market shares of the news outlets with more coverage of sensitive news, and an increase in the amount of Ukraine-crisis news is more likely to lead to an increase in the market shares of the news outlets with less pro-government slant and higher volume of slant.

However, we cannot conclude that the consumers prefer the news outlets with more reporting about sensitive news and less pro-government slant in the Ukraine crisis coverage. There are multiple alternative explanations for the observed relationship between the news outlets consumption changes and the amount of sensitive news in the market. First, consumer preferences are likely to have some degree of heterogeneity, and preference heterogeneity can be responsible for the observed patterns in the market shares changes. For example, if some consumers prefer the pro-government slant in the Ukraine-crisis coverage and others prefer the anti-government slant, the market shares of the news outlets with high volume of slant will increase, but due to the consumption from two separate consumer segments and not due to the preference for a higher volume of slanted news. Another example is the sorting of consumers who prefer the independent news outlets to reading the news on the days with a lot of internally sensitive news due to the positive correlation in these consumer preferences. Such sorting will lead to higher market shares of the less pro-government biased news outlets on the days with more sensitive news, while an average consumer might prefer more pro-

<table>
<thead>
<tr>
<th>( F_{j}^{Ukr} )</th>
<th>( \hat{b}_{j}^{Ukr} )</th>
<th>( \hat{b}_{j}^{Ukr} )</th>
<th>( \hat{b}_{j}^{Ukr} )</th>
<th>( \hat{b}_{j}^{Ukr} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_{j}^{-} )</td>
<td>( -0.010 )</td>
<td>( -0.010 )</td>
<td>( 0.055 )</td>
<td>( 0.035 )</td>
</tr>
<tr>
<td>( V_{j}^{+} )</td>
<td>( 0.022 )</td>
<td>( 0.021 )</td>
<td>( -0.021 )</td>
<td>( 0.024 )</td>
</tr>
</tbody>
</table>

Controls (from the regression 1):

- Weekday FE: \( N \) \( Y \) \( Y \) \( Y \)
- Time trend polynomial (4-order): \( N \) \( N \) \( Y \) \( N \)
- Week FE: \( N \) \( N \) \( N \) \( Y \)

Controls are included in the regression (1) estimating \( \hat{b}_{j}^{Ukr} \). Standard errors are heteroskedasticity consistent.
government biased coverage. Second, an increase in the market share of the news outlets with a high volume of slant can be driven by the conscientious consumers, who will read the news outlets with the extreme ideological positions on days with more sensitive news. In order to account for these alternative explanations, in the next section we formulate a structural demand model and use the individual-level consumption data to estimate consumer preferences.

5 Model

5.1 Nature

There is a set of possible news events $S$. Each event is related to one of three news topics: non-sensitive news for the government, internally sensitive news, and news about the Ukraine crisis. Every day $t$, nature produces news about a subset of these events, $S_t$. We denote the number of events that happens about a given topic as $N^x_t$, where $x$ corresponds to non-sensitive, internally sensitive or Ukraine-crisis news, $x = \{\text{Non, IS, Ukr}\}$. We assume that the news-market participants take the underlying production process as a given.

As researchers, we do not observe $N^x_t$. Instead, we measure the relative importance of topic $x$ on day $t$ with a fraction of news articles about topic $x$ in the market (across all news outlets), $F^x_t = \frac{\sum_j N^x_{jt}}{\sum_x \sum_j N^x_{jt}}$, where $N^x_{jt}$ is the number of articles about topic $x$ by the news outlet $j$ on day $t$.

5.2 News Outlets Reporting and Government Control

The market contains $J$ news outlets. Each news outlet $j$ is given its type $\text{type}_j$: independent, influenced, GC, Ukrainian, or international. News outlet $j$ chooses its quality $\alpha_j$, its level of reporting about the internally sensitive and Ukraine-crisis news, $\bar{F}^\text{IS}_j$ and $\bar{F}^\text{Ukr}_j$, and its valence and volume of slant in the Ukraine-crisis news coverage, $V^{-}_j$ and $V^{+}_j$.\footnote{Quality $\alpha_j$ represents an average persistent preference of the outlet $j$ across consumers in the market.} For simplicity, we assume that the news outlets make their choices only once and commit to the same quality and ideological positions throughout the sample period.

Government control affects the ideological positions of the news outlet. The censorship constraint affects the reporting of news outlets. Under censorship, the government determines which fraction of sensitive news is reported by the GC news outlets. The propaganda constraint affects the valence and volume of slant reported by the GC news outlets.
5.3 Demand

There are $I$ consumers in the market. We assume that consumers are in the market for online news on the days when they are browsing online. On each consumption occasion $\tau$ on day $t$, consumer $i$ can choose one news outlet, or choose an outside option of not consuming any news.\footnote{Following Gentzkow and Shapiro (2015), we restrict consumer choice to at most one news outlet per consumption occasion because it is impractical for people to read multiple news articles at the same time. Our set-up does not restrict consumers to navigate to multiple news outlets on the same day $t$.} As before, we define a news consumption of an outlet $j$ as navigation to at least one news article on the outlet’s $j$ website by consumer $i$ on day $t$. Thus, consumer can visit news outlet $j$ on day $t$ at most once. The sequence of consumption occasions $\tau$ is determined by the earliest news article visited by the consumer $i$ on an outlet $j$ on day $t$.

We assume that consumers have information about the relative importance of news topics over time, $F_{t}^{IS}$ and $F_{t}^{Ukr}$ $\forall t$. Given that we define the news consumption as navigation to the news article, consumers have an opportunity to acquire the knowledge about $F_{t}^{IS}$ and $F_{t}^{Ukr}$ beforehand, for example, from the information on the search engines or even a list of news articles on the main page of a particular website. In addition, we assume that consumers know the ideological positions of the news outlets, $\bar{F}_{J}^{IS}$, $\bar{F}_{J}^{Ukr}$, $V^{-}$ and $V^{+} \forall j$.

The news preferences of consumer $i$ are defined over four dimensions: fixed preferences for the news outlets, $\alpha_{i} = \{\alpha_{i1}, \ldots, \alpha_{iJ}\}$, preferences for the reporting of a news topic $x$, $\beta_{i}^{x}$, preferences for the valence and volume of the ideological slant, $\gamma_{i} = \{\gamma_{i}^{-}, \gamma_{i}^{+}\}$, and preference for the ideological diversity of the sensitive news, $\rho_{i}$. We assume that the preferences of consumer $i$ are fixed over time.

On each day $t$, consumer $i$ can have multiple news consumptions $T_{it} = \{1, \ldots, M + 1\}$, where $M$ is the total number of news outlets in the market. That is, at each day $t$, consumer $i$ chooses the outlets sequentially on occasions $\tau = \{1, \ldots, T_{it}\}$, where in the last choice occasion $T_{it}$ he chooses an outside option. For simplicity, we treat the number of choice occasions $T_{it}$ as exogenous.\footnote{This assumption limits our ability to simulate changes in the number of choice occasions in the counterfactual scenarios, but still allows us to estimate consumer preferences for the variety in the news outlets’ ideology.} We define the utility of consumer $i$ from an outlet $j$ on day $t$ and on consumption occasion $\tau$ as

$$u_{ijt\tau} = \alpha_{ij} + F_{t}^{Ukr} (\eta_{i}^{Ukr} + F_{j}^{Ukr} \beta_{i}^{Ukr} + V_{j}^{-} \gamma_{i}^{-} + V_{j}^{+} \gamma_{i}^{+} + |V_{j}^{+} - V_{y_{\tau-1}}^{+}| (\tau > 1)\rho_{i}) +$$

$$+ F_{t}^{IS} (\eta_{i}^{IS} + F_{j}^{IS} \beta_{i}^{IS}) + |V_{j}^{-} - V_{y_{\tau-1}}^{-}| (\tau > 1)\eta_{i}^{-} + s_{it} \eta_{i}^{s} + \epsilon_{ijt\tau},$$

where $\epsilon_{ijt\tau}$ is an idiosyncratic shock to consumer utility, and $\eta_{i} = \{\eta_{i}^{Ukr}, \eta_{i}^{IS}, \eta_{i}^{-}, \eta_{i}^{s}\}$ is a set.
of reduced-form parameters. Coefficients $\eta^U_k$ and $\eta^I_k$ explain the changes in the consumer utility due to the changes in the amount of news topics happening on day $t$. The measure $|V^j - V^y_{\tau-1}|$ captures the ideological distance in the Ukraine crisis slant between consumer $i$’s current and previous consumption choice (denoted as $y_{\tau-1}$) on day $t$, and it does not affect consumer utility on the first consumption occasion on day $t$. Thus, coefficient $\eta^-$ captures the baseline variety-seeking behavior of consumers and allows the interpretation of $\rho_i$ as the preference for ideological diversity in the Ukraine-crisis coverage. Finally, state variable $s_{i\tau}$ is an indicator variable equal to 1 if an outlet $j$ was consumed on day $t$ on one of the previous choice occasions $1, \ldots, \tau - 1$. Given that we allow each news outlet to be consumed only once on day $t$, $s_{i\tau}$ allows us to capture the fact that consumers do not return to an outlet $j$ after the consumption. The choice of an outside option is defined as being online but not navigating to the news articles, and it is normalized to $u^{0\tau} = \epsilon^{0\tau}$.

We note that there are multiple implicit assumptions underlying the specified model. First, we treat the number of choice occasions $T_{it}$ as exogenous. This assumption limits our ability to simulate changes in the number of choice occasions in the counterfactual scenarios, but still allows us to estimate consumer preferences for the variety in the news outlets’ ideology. Second, we assume that quality is defined on the outlet level and does not differ across the news topics. Third, we assume that consumers do not incur switching costs when making day-to-day outlet consumption decisions. Finally, we assume that consumers’ preferences for topics do not vary with the relative importance of the topics.

### 5.4 Identification

Identification of consumer preferences relies on the exogenous shifts in the amount of sensitive news over time and the reporting and ideological positions of the news outlet. On the days with little sensitive news ($F^U_t = F^I_t = 0$), the news consumption utility for consumer $i$ comes from the fixed preferences of this consumer from the new outlets, $\alpha_i$. Thus, consumption choices on these days identify $\alpha_i$. Reduced-form parameters $\eta^-$ and $\eta^s$ are identified from the occasions with multiple news outlets consumed within a day. On the days with more sensitive news ($F^U_t > 0; F^I_t > 0$), consumer $i$ derives utility both from the fixed effects $\alpha_i$ but also from her preferences for the sensitive news topics. Parameters $\eta^U_k$ and $\eta^I_k$ are identified from the likelihood to consume any news outlet on the days with more sensitive news. Ideological preferences $\beta^I_i, \beta^U_i, \gamma^-, \gamma^+$ are identified from consumer preferences for topics do not vary with the relative importance of the topics.

---

53 Given that we observe only relative importance of news topics over time, $F^U_t$ and $F^I_t$, we normalize consumer preference for non-sensitive news to zero, $\eta^{Non}_i = \beta^{Non}_i = 0$. 

32
i’s switching on the days with more sensitive news. Finally, preference for the ideological diversity, $\rho_i$, is identified from the changes in the ideological “variety-seeking” behavior of consumers on the days with more news about the Ukraine crisis.

6 Estimation and Results

6.1 Consumer Sample

The demand model specified above makes several assumptions about the information that consumers have when navigating to the news articles. In particular, it assumes that consumers know the average ideological positions of the news outlets and the relative importance of the sensitive news topics on a given day. These assumptions are more likely to hold for news consumers who read news on multiple days over the sample period. Thus, for the demand estimation we focus on the news readers who consume news at least 10 days in our sample period. This corresponds to a sample of the 52,568 news consumers.\(^{54}\) While this sample corresponds only to 24.5% of news readers in the market, these consumers account for 92.2% of all the news articles read in the sample period. For the demand estimation, we also focus on the top 36 online news outlets in the sample, due to low market shares of the rest of the news outlets.

News readers in the selected sample have 4,456,161 consumption occasions. On the majority (63.9%) of the consumption days, news readers in the selected sample visit only 1 news outlet. However, conditional on visiting more than one news outlet on day $t$, news readers navigate to an average of 2.84 news outlets, and on 7 consumption occasions consumers visit more than 20 news outlets in the sample.

6.2 Estimation

We estimate the distribution of $\theta_i = \{\alpha_{ij}, \eta_i \text{Ukr}, \eta_i \text{IS}, \eta_i \text{u}, \beta_i \text{Ukr}, \beta_i \text{IS}, \gamma_i^-, \gamma_i^+, \rho_i\}$ using a Bayesian hierarchical model. We assume that $\epsilon_{ijt\tau}$ follows a type-1 extreme value distribution. At each choice occasion $\tau$ on day $t$, a consumer chooses an outlet $j$ such that $u_{ijt\tau} \geq u_{ij't\tau} \forall j' \in \{0, \ldots, J\} : j' \neq j$.

Denote consumers’ choices as $y$. The probability that consumer $i$ chooses news outlet $j$\(^{54}\) Out of 214,375 news consumers who visit a news article page at least once over the sample period.
at on day $t$ on the consumption occasion $\tau$ is

$$\pi(y_{it\tau} = j | \theta_i) = \frac{\exp(u_{ijt\tau}(\theta_i))}{1 + \sum_{j'} \exp(u_{ij'\tau}(\theta_i))}.$$

The likelihood of $\theta_i$ observing a sequence of choices $y_i$ is

$$L(\theta_i | y_i) = \prod_t \prod_\tau \prod_j \pi(y_{it\tau} = j | \theta_i)^{I(y_{it\tau} = j)}.$$

The first-stage prior on $\theta_i$ is a normal distribution, with the normal prior over its mean and the inverse Wishart prior over the covariance matrix:

$$\theta_i \sim N(\mu, \Sigma),$$

$$\mu \sim N(\bar{\mu}, \Sigma \otimes a_{\mu}^{-1}),$$

$$\Sigma \sim IW(\nu_{\Sigma}, V_{\Sigma}).$$

We estimate the distribution of the parameters $\theta$ by simulating from the posterior distribution using an MCMC hybrid sampler. Appendix 10.10 provides the computational details.

### 6.3 Estimation Results

Tables 9 and 10 present the posterior point estimates of consumer preferences. Table 9 describes the structure of consumer preferences for news coverage. First, an average consumer prefers the news outlets with more coverage of the internally sensitive news ($E(\hat{\beta}^{IS}) = 0.021$) and the Ukraine-crisis news ($E(\hat{\beta}^{Ukr}) = 0.101$). However, there is substantial heterogeneity in consumer preferences, with 42.11% of consumers having a preference for non-sensitive news over the internally sensitive news and 32.4% of consumers having a preference for non-sensitive news over the Ukraine-crisis news. Second, an average consumer prefers the Ukraine-crisis news coverage with less pro-government slant ($E(\hat{\gamma}^-) = 0.068$) and less slant in general ($E(\hat{\gamma}^+) = -0.01$), but once again there is significant heterogeneity in consumer preferences, with 38.54% of consumers having a preference for more pro-government slant. Finally, the vast majority of consumers in the sample, 74.5%, prefer to read more ideologically-similar news on the days with more sensitive news, suggesting that only a minority of consumers in the sample are conscientious consumers.

Table 10 focuses on the persistent consumer preferences for news outlets, $\alpha_j$, representing the stable characteristics of the news outlets, such as quality. To compare these preferences
Table 9: Posterior point estimates of consumer preferences for news coverage.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
<th>% of users &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \eta^{IS} )</td>
<td>0.001</td>
<td>0.135</td>
<td>50.41</td>
</tr>
<tr>
<td>( \beta^{IS} )</td>
<td>0.021</td>
<td>0.113</td>
<td>57.89</td>
</tr>
<tr>
<td>( \eta^{Ukr} )</td>
<td>0.262</td>
<td>0.568</td>
<td>68.56</td>
</tr>
<tr>
<td>( \beta^{Ukr} )</td>
<td>0.101</td>
<td>0.222</td>
<td>67.60</td>
</tr>
<tr>
<td>( \gamma^- )</td>
<td>0.068</td>
<td>0.235</td>
<td>61.46</td>
</tr>
<tr>
<td>( \gamma^+ )</td>
<td>-0.010</td>
<td>0.166</td>
<td>47.56</td>
</tr>
<tr>
<td>( \eta^- )</td>
<td>0.152</td>
<td>0.345</td>
<td>66.19</td>
</tr>
<tr>
<td>( \rho )</td>
<td>-0.109</td>
<td>0.164</td>
<td>25.50</td>
</tr>
</tbody>
</table>

The posterior standard deviation estimates are in the brackets.

across the news outlet types, we aggregate \( \alpha_j \) by outlet types, \( \hat{\alpha}_{\text{type}} \), and demean it by average \( \alpha_j \) across all the news outlets, \( \hat{\alpha} \). An average consumer prefers the GC news outlets the most (\( \hat{\alpha}_{GC} \) – \( \hat{\alpha} \) = 0.606), followed by the influenced (\( \hat{\alpha}_{Inf} \) – \( \hat{\alpha} \) = 0.312) and independent (\( \hat{\alpha}_{Inf} \) – \( \hat{\alpha} \) = −0.072) news outlets. While there is substantial heterogeneity in consumer preferences, the vast majority of consumers prefer the quality of the GC and influenced news outlets to the quality of an average news outlet (88.5% and 82.48%, respectively).

Results in Tables 9 and 10 reveal a nuanced picture. On the one hand, results suggest that quality of the GC news outlets is the primary driver of their demand. First, we find that the majority of consumers prefer the quality of the GC news outlets over the average news outlet (88.04%) and over the average independent news outlet (78.4%). Second, the majority of consumers get disutility from the ideological positions of the GC news outlets, with 57.89% of consumers preferring more coverage about the internally sensitive news and 61.46% preferring less pro-government slanted news about the Ukraine crisis. On the other hand, some consumers prefer the pro-government slant in the Ukraine-crisis news, and they might navigate to the GC news outlets because of it.

To understand the demand for the GC news outlets, we compare the magnitudes of

\[0.3\% \text{ posterior standard deviation.}\]
Table 10: Posterior point estimates of persistent preferences for news outlets.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D.</th>
<th>% of users &gt; 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}$</td>
<td>-6.681</td>
<td>1.273</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha}_{GC} - \hat{\alpha}$</td>
<td>0.606</td>
<td>0.512</td>
<td>88.50</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.27)</td>
</tr>
<tr>
<td>$\hat{\alpha}_{Ind} - \hat{\alpha}$</td>
<td>-0.072</td>
<td>0.618</td>
<td>45.32</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.44)</td>
</tr>
<tr>
<td>$\hat{\alpha}_{Inf} - \hat{\alpha}$</td>
<td>0.312</td>
<td>0.344</td>
<td>82.48</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.37)</td>
</tr>
<tr>
<td>$\hat{\alpha}_{Int} - \hat{\alpha}$</td>
<td>-1.138</td>
<td>1.263</td>
<td>17.41</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.008)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>$\hat{\alpha}_{Ukr} - \hat{\alpha}$</td>
<td>-2.678</td>
<td>2.297</td>
<td>10.58</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.033)</td>
<td>(0.21)</td>
</tr>
</tbody>
</table>

The posterior standard deviation estimates are in the brackets.

consumer preferences for news coverage and their persistent preferences. Recall that we have normalized the measures of share of sensitive news on day $t$, $F_{IS}^t$ and $F_{Ukr}^t$, to have a unit mean, and the reporting and slant decisions of the news outlets, $\bar{F}_{IS}^t$, $\bar{F}_{Ukr}^t$, $V_j^-$ and $V_j^+$, to have a zero mean and a unit standard deviation. This way, we can interpret $\hat{\eta}_{IS}$ and $\hat{\eta}_{Ukr}$ estimates as a difference in the utility from reading a news outlet with an average reporting on sensitive news on a day with an average amount of sensitive news and a day with no sensitive news, and $\hat{\beta}_{IS}$, $\hat{\beta}_{Ukr}$, $\hat{\gamma}^-$ and $\hat{\gamma}^+$ estimates as a utility from one standard deviation more reporting on sensitive news by this outlet, one standard deviation less pro-government slant, and one standard deviation more slant in general. For example, an average consumer gets 0.262 more utility from reading an average news outlet on a day with an average amount of the Ukraine-crisis news compared to a day with no Ukraine-crisis news and another 0.068 of utility from an outlet with one standard deviation less of pro-government slant.

Using these calculations, we compare the utility consumers get from the ideological positions of the GC and independent news outlets. For internally sensitive news, our classification suggests that censorship is the primary mechanism of government control, so we compare the utilities consumers get from the amount of coverage about internally sensitive news on day $t$, $F_{IS}^t \hat{\beta}_{IS}(F_{IS}^{GC} - F_{IS}^{Ind})$. The difference in coverage between the GC and independent news outlets is 2.17 standard deviations, so an average consumer gets $0.021 \times 2.17 = 0.046$ more utility from an average independent news outlet on days with an average amount of sensitive news compared to days with no sensitive news. This utility difference is small in comparison to the utility difference of 0.678 between the GC and independent news outlets.
Subfigure 11 (a) plots changes in the difference of consumer utilities from an average GC and independent news outlet as the volume of internally sensitive news increase in the market. The fraction of consumers who prefer an average GC news outlet to the independent news outlet stays almost the same, reducing from 78.4% on days with no sensitive news to 73.8% on days with twice the average of sensitive news, showing that the quality difference is more important for consumers than the ideological difference.\textsuperscript{56}

Figure 11: Distribution in the expected utility difference between an average GC and independent news outlet.

For the Ukraine-crisis news, our results suggest that slant is the primary mechanism of government control, so we compare the utilities consumers get from slant in the Ukraine-crisis coverage on day $t$, $F_{t}^{Ukr}(\hat{\gamma}^{-}(V_{GC}^{-} - V_{Ind}^{-}) + \hat{\gamma}^{-}(V_{GC}^{+} - V_{Ind}^{+}))$. An average GC news outlet has 2.45 standard deviations more pro-government slant ($V_{GC}^{-} - V_{Ind}^{-} = -2.45$) and 1.12 standard deviation more slant in general ($V_{GC}^{+} - V_{Ind}^{+} = 1.12$) compared to an average independent news outlet, so an average consumer gets $0.068 \times 2.45 + (-0.01) \times (-1.12) = 0.178$ less utility from an average GC news outlet on days with an average amount of Ukraine-crisis news compared to days with no Ukraine-crisis news. Thus, while Ukraine-crisis news coverage plays a more important role in consumers’ utilities than the coverage of internally sensitive news, it is still lower than the role of quality. Subfigure 11 (b) plots changes in

\textsuperscript{56}Posterior standard deviation estimate of the difference is 0.26%.
the difference of consumer utilities from an average GC and independent news outlet as the volume of the Ukraine-crisis news increases in the market. With an increase in the volume of the Ukraine-crisis news, the share of consumers who prefer the GC news outlets falls, with 60.7% (0.34%) of consumers having a preference for the GC news outlet on days with twice the average of the Ukraine-crisis news. Still, the majority of consumers prefer an average GC news outlet to an average independent news outlet, emphasizing the importance of the quality difference between the GC and independent outlets.

The results above show that the quality difference between the GC and independent news outlets plays a more important role in the demand for the GC news outlets than the ideological differences. In section 7 below, we further examine the relative importance of the quality and ideological positions by simulating the market shares of the news outlets under different levels of quality and government control. Before we move to the counterfactuals, we discuss two additional characteristics of the consumer preference estimates.

First, given that we find that persistent preferences of consumers play a crucial role in the demand for news, we describe the nature of the persistent preferences of consumers, $\alpha_{ij}$. While we refer to the persistent preferences of consumers as quality, they potentially include any characteristics of the website, such as the breadth of news coverage and brand capital. In particular, the ideological position of the news outlet might affect persistent preferences of the consumers in the long run if consumers accumulate some brand capital by navigating to the news outlet. To check if persistent preferences of consumers are correlated with their ideological positions, in the Appendix 9 we analyze the correlation structure between the persistent preferences of consumers for the news outlets and examine if their correlations are related to the ideological positions of these news outlets. We find that the correlations in persistent preferences are higher for the news outlets that are more ideologically-similar, suggesting that ideology plays a role in brand capital formation. Thus, we need to interpret estimates of $\beta$ and $\gamma$ as a short-term effect of the ideological positions on the market shares.

Second, structural demand estimates allow us to separate out alternative mechanisms behind the demand for biased news coverage: confirmation bias, conscientious consumption of news, and entertainment. Using the results in Table 9, we conclude that few consumers in this market behave like the conscientious news readers (Mullainathan and Shleifer, 2005); only 25.5% of consumers start navigating to the ideologically-diverse news outlets on days with more sensitive news. Results also show that a substantial share (47.56%) of the news readers care simply about the volume of slant in the news, which suggests that slanted news have an entertainment value for these consumers. However, the data shows that confirmation
bias is the most likely explanation for the demand of the consumers for the ideologically slanted content, with the vast majority of people navigating to the more ideologically-similar news outlets on the days with a lot of sensitive news.

7 Counterfactuals

Estimation results have revealed that the quality difference between the GC and independent news outlets plays a more important role in the demand for the GC news outlets than their ideological position. But how much market share do the GC news outlets gain because of their superior quality, and how much market share do they lose due to the pro-government bias? In this section, we address these questions by simulating the market shares of the news outlets under different levels of quality and government control. Due to the nature of our estimates, we focus on a short-term effect of a change in quality and ideology, with changes in the ideological positions affecting the consumers only through their preferences for the news coverage, $\beta$ and $\gamma$, and not through the potential changes in the persistent preferences.\footnote{Thus, we do not consider long-term factors like brand capital formation.} Throughout the counterfactuals, we treat the average ideological position of the independent news outlets as “unbiased,” and deviations from this average position as a result of government control.\footnote{Naturally, such an approach does not account for the product differentiation in the ideological space. An alternative empirical strategy is to specify the supply-side model with some outlets being under a government constraint, to estimate the costs parameters and to examine the counterfactual decisions of the firms after removing the constraint. Given that we focus on the short-term effects of the government control, we leave this for future work.} In order to speed up the counterfactual simulation, we approximate the news realizations $F_{t}^{IS}$ and $F_{t}^{Ukr}$ by the centers of 20 clusters of these variables and simulate one choice occasion per consumer per day.\footnote{Standard k-means clustering algorithm is applied to cluster the observed $F_{t}^{IS}$ and $F_{t}^{Ukr}$.}

Table 11 presents the simulated market shares under different levels of government control and quality of the GC news outlets. Columns (1) and (2) compare the predicted market shares under the current quality of the GC news outlets and under the average quality of the independent news outlets $\alpha_{j}^{\text{low}} = \alpha_{j} - \frac{\sum_{j' \in GC} \alpha_{j'}}{\sum_{j' \in GC} 1} + \frac{\sum_{j' \in Ind} \alpha_{j'}}{\sum_{j' \in Ind} 1} \forall j \in GC$. Under the lower quality regime, the market share of the GC news outlets decreases by 44.6%, from a 7.19% share to a 3.98% share. The influenced news outlets benefit the most from this reduction in quality as their market share increases by 5.8%. However, most of the switching consumers, 72.3%, choose not to read online news outlet after the quality decreases.

Columns (3)-(6) present the predicted market shares under the counterfactual levels of quality and government control.
Table 11: Simulated market shares for different levels of government control and quality of the GC news outlets.

<table>
<thead>
<tr>
<th>Outlet Type</th>
<th>(1) Actual</th>
<th>(2) Low $\alpha_{GC}$</th>
<th>(3) Direct No control</th>
<th>(4) Indirect No control</th>
<th>(5) Both No control</th>
<th>(6) More control</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC ($sh_{Gov}$)</td>
<td>7.19</td>
<td>3.98</td>
<td>8.64</td>
<td>7.11</td>
<td>8.47</td>
<td>7.23</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.032)</td>
<td>(0.007)</td>
<td>(0.030)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Influenced ($sh_{Inf}$)</td>
<td>9.91</td>
<td>10.49</td>
<td>9.66</td>
<td>10.60</td>
<td>10.27</td>
<td>9.97</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.019)</td>
<td>(0.017)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Independent ($sh_{Ind}$)</td>
<td>6.35</td>
<td>6.60</td>
<td>6.19</td>
<td>6.25</td>
<td>6.12</td>
<td>5.92</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>International ($sh_{Ind}$)</td>
<td>0.63</td>
<td>0.66</td>
<td>0.61</td>
<td>0.62</td>
<td>0.60</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Ukrainian ($sh_{Ukr}$)</td>
<td>0.97</td>
<td>1.00</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>None above ($sh_{Outside}$)</td>
<td>74.95</td>
<td>77.27</td>
<td>73.94</td>
<td>74.46</td>
<td>73.59</td>
<td>75.28</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.033)</td>
<td>(0.024)</td>
<td>(0.045)</td>
<td>(0.016)</td>
</tr>
</tbody>
</table>

The market share are in percent of the entire market. The posterior standard deviation estimates are in the brackets.

government control. We distinguish between direct control executed through ownership, as in the case of GC news outlets, and indirect control executed through the influence of news outlet owners, as in the case of influenced news outlets (Gehlbach and Sonin, 2014).

Column (3) presents the results for a market with no direct control, a scenario when the GC news outlets have average ideological positions similar to the independent news outlets, 

$$F_{j}^{IS,new} = F_{j}^{IS} - \frac{\sum_{j' \in GC} F_{j'}^{IS}}{\sum_{j' \in GC} 1} \frac{1}{1 + \sum_{j' \in Ind} \frac{V_{j'}^{-}}{\sum_{j' \in Ind} 1}} \frac{1}{1 + \sum_{j' \in Ind} \frac{V_{j'}^{+}}{\sum_{j' \in Ind} 1}} \forall j \in GC.$$  

Without the direct control, market shares of the GC news outlets increase by 20.2%, with most of the traffic coming from the extensive margin. Thus, we confirm that direct control is a binding constraint on the GC news outlets as they are losing the market shares because of the pro-government bias.

What do the simulated changes in the market share of the GC news outlets imply about their profitability? While we do not have detailed information about the revenue sources of the news outlets, we can do a simply back-of-the-envelope calculation based on the advertising market size. For the online news outlets in Russia in 2013-2015, the main source of revenue is display advertising.\(^{60}\) In 2014, the total expenditure on display advertising on the Russian internet was 19.1 billion rubles\(^ {61}\), which is around $318 million using the

\(^{60}\)Only one of the news outlets in the sample, slon.ru, used paid subscription starting in 2015.

\(^{61}\)http://www.akarussia.ru/knowledge/market_size
exchange rate of the end of 2014 of 60 rubles for a dollar. Even if we assume that the online news market gets all the display advertising revenues, the 1.45 percentage points reduction in the market share of the GC news outlets due to direct control is small, corresponding to $18.41 million dollars. For comparison, government subsidies to mass media in Russia in 2015 were 72.6 billion rubles ($1.21 billion), which is around 65.7 times higher than the advertising loss.\footnote{Source: http://www.rbc.ru/politics/29/06/2015/55912ffa9a7947453982cda9. The same exchange rate is used. The total of 72.6 billions rubles includes subsidies to the television and print media.}

We discuss other control scenarios in columns (4)-(6). In column (4), we examine the market shares under a regime with no indirect control, with the ideological positions of the influenced news outlets adjusted to match the independent news outlets. Similar to the direct control, indirect control is a binding constraint on the affected outlets, with influenced outlets losing 6.5% of their market share. Results in column (5) show that under no control (direct or indirect) GC and influenced news outlets both gain in the market shares, although these changes are smaller compared to the cases with only direct or indirect control.\footnote{These results suggest that direct and indirect control are complementary from a perspective of a government: GC and influenced news outlets have higher readership when both direct and indirect control is imposed (7.19 + 9.91 = 17.1%) compared to the regimes with only direct and only indirect control (7.11 + 9.66 = 16.77%).}

In column (6) we examine the scenario under which all independent news outlets become indirectly controlled, a feasible scenario based on the events of 2016-2017.\footnote{By the middle of 2016, several independent news outlets had to change their ownership due to a new law (https://rg.ru/2016/01/01/smi-site-anons.html), and rbc, one of the top online news outlets in Russia, had to change the editorial team due to the government pressure (http://www.bbc.com/russian/news/2016/05/160513_rbc_badanin) as well as its ownership later in 2017 (http://www.forbes.ru/milliardery/346333-berezkin-kupil-u-prohorova-rbk).} Independent news outlets lose 6.8% of their market share if their average ideological position is changed to match the ideological position of the influenced news outlets. Using the back-of-the-envelope calculations similar to the above, this market share loss corresponds to an upper bound of $5.46 million, implying that it would not be expensive for the government to convince the independent news outlets to become influenced if independent outlets cared only about the advertising revenues.

### 7.1 Online Media Power of the Government

We have shown that the GC news outlets are able to maintain a higher market share in the online market partly because of their superior quality. How much does this high level of quality or brand capital help the GC news outlets to increase their media power? Following
Prat (2017), we focus on the share of attention that consumers pay to each news outlet. Unlike Kennedy and Prat (2017), we do not observe the consumption of consumers on other platforms, such as TV and print, so we focus on the online attention of the news consumers. Using the demand model, we extend the definition of the attention share of consumer $i$ on day $t$ to an outlet $j$ as

$$\Pr(y_{it} = j) / (1 - \Pr(y_{it} = 0)),$$

where $0$ is an outside option. Aggregating this across days and consumers, we get the attention share of an outlet $j$

$$E_{i,j} (\Pr(y_{it} = j) / (1 - \Pr(y_{it} = 0))).$$

Using this definition, the attention share of the GC news outlets is 33.4% (0.1%), corresponding to the media power of 0.5 under the worst-case scenario assumptions, meaning that the government is able to swing 25-75% elections into a draw.\(^{65}\)

To understand the role of the GC news outlets’ quality in their media power, we compute the attention shares of consumers under the lower quality of the GC news outlets, as in the case of column (2) of Table 11. Under this quality, the online attention share of the GC news outlets reduce by 11.12 percentage points to 21.78% (0.06%), corresponding to 0.28 media power. Such media power allows the government to swing 36-64% elections in to a draw. Thus, around 1/3 of the attention share of the GC news outlets and almost half of their media power is driven by the high quality or brand capital of the GC news outlets, which we refer to as “brand media power.”

In addition to the overall attention share of the news outlets, demand estimates allow us to study the attention share of the GC news outlets over consumers who have a distaste for the pro-government bias:

$$\Pr(y_{it} = j | \Delta U^x_i < 0) / (1 - \Pr(y_{it} = 0 | \Delta U^x_i < 0)),$$

where $\Delta U^x_i$ is the utility consumer $i$ gets from the pro-government bias in sensitive news $x$ topic, $\Delta U^IS_i = \beta^IS_i (\bar{F}^IS_{GC} - \bar{F}^IS_{Ind})$ and $\Delta U^Ukr_i = \gamma^- (V^-_{GC} - V^-_{Ind}) + \gamma^- (V^+_{GC} - V^+_{Ind})$. We use this measure to compute the attention share of the GC news outlets over consumers with $\Delta U^x_i < 0$ on the days with a lot of sensitive news, a case where the GC news outlets can successfully prevent a motivated consumer from learning the information. The attention

---

\(^{65}\)The worst-case scenario includes the assumption that the readers are naive and do not understand that the GC news outlets are trying to persuade them. For more details, see Prat (2017) and Kennedy and Prat (2017).
shares are 31.1% for a big internally sensitive news day and 21.5% for a big Ukraine-crisis news day. Under the lower quality of the GC news outlets, the attention shares on such days change to 19.9% and 13.1%, respectively. Thus, the high quality of the GC news outlets allows them to capture an additional 8.4-11.2 percentage points of consumers who prefer the ideological coverage of the independent news outlets.

8 Conclusion

In the new era of broad access to information, it is critical to understand whether and how governments can control public opinion online. In this paper, we show that the governments can successfully control the news outlets and have high media power even in the presence of the independent news outlets. Using an example of the online news market in Russia, we show that the government-controlled news outlets maintain a substantial market share even though the majority of the population dislikes their ideological bias. The main driver for these results are the persistent preferences of consumers for the controlled news outlets, reflecting their characteristics such as quality and brand capital. High “quality” is responsible for 44.6% of the controlled news outlets’ market shares and one-third of their attention share.

In contrasts, removing the ideological bias from the coverage of the controlled news outlets would increase their market shares only by 20.2%, corresponding to a rough estimate of $18.41 million in advertising revenues. This implies that it is relatively cheap for the government to cover the news outlets’ losses from the pro-government bias in the news. For example, Russian government needs to subsidize the independent news outlets only $5.46 million a year to cover their potential advertising losses from starting to report like the government-influenced news outlets.

Finally, the behavior of the majority of the news readers in Russian online news market is consistent with the confirmation bias theory, with only 25.5% of consumers behaving like “conscientious” news readers and starting to read more ideologically-diverse news on the days with a lot of sensitive events.

We note, however, that our analysis of the effect of government control on the market share and media power of the GC news outlets is limited to the short-term effects. First, throughout the work we focus on the formed preferences of consumers, ignoring the potential changes induced by persuasion of the news outlets’ coverage, or simply preference changes over time. Such changes might increase or decrease the role of the ideological position of

\footnote{A big sensitive news day is a day with three times the average amount of sensitive news.}
the GC news outlets in the long-run. Second, ideological positions of the news outlets might affect the formation of persistent preferences for these news outlets, and in the long run, changes in the ideological positions will also have an impact on the brand capital of the outlets. Third, our measure of the ideological bias of the GC news outlets is based on the comparison of the GC and independent news outlets, ignoring potential self-censorship of the independent outlets. Changes in the level of government control might change the degree of self-censorship. Finally, once government control is removed, changes in the ideological positions of the formerly-controlled news outlets will trigger a supply-side response from the other news outlets in the market, which might lead to some news outlets introducing pro-government biased coverage to fit to the preferences of a minority of consumers who prefer such coverage. The question of the long-term effect of the government control is an interesting area of future research.

References


9 Appendix A: Correlation in Persistent Preferences

Do higher $\alpha$ estimates represent a higher quality of the GC news outlets, a result of a government’s investments? Or is there some outlet-specific accumulated brand capital, which might be driven by the ideological positions of the news outlets? While we do not model brand capital formation, we can examine the correlation in the persistent brand preferences, $\alpha_{ij}$, across the news outlets. If $\alpha_{ij}$ estimates are driven primarily by the ideological position of the news outlet, consumer persistent preference estimates should be highly correlated across the news outlets with the same ideological position. In contrast, if $\alpha_{ij}$ estimates are driven primarily by the quality of the news outlets, correlation in persistent preference should be driven by the overall quality of the news outlets, $\bar{\alpha}_j$.

Figure 12 summarizes the estimates of correlation in persistent outlet preferences, $\alpha_{ij}$, across the news outlets. Similar to Table 10, we subtract the average preference for news outlets, $\bar{\alpha}_i$, from the $\alpha_{ij}$ to exclude the influence of consumer $i$’s preference for news in general. News outlets are colored by their types, correspond to the legend in Figure 9 and are sorted by the degree of correlation between each other. The results suggest that there is at least some correlation in consumer persistent preferences driven by the news outlets’ ideology. For example, consumer preferences for all Ukrainian and international news outlets are highly positively correlated among each other and are negatively correlated with the GC news outlets. Visually, we can also conclude that news outlets are grouped by their type. For example, independent news outlets tend to be highly correlated with other independent news outlets, and so are the GC news outlets. At the same time, news outlets are not perfectly grouped by types, suggesting that quality might also play a role in persistent preferences.

To test the alternative explanations for persistent preferences of consumers more formally, we regress the estimated correlations on the ideological and quality distance between the news outlets. We measure the distance as the absolute value in the news outlets characteristics, such as the amount of reporting about sensitive news, $F^{IS}_j$ and $F^{Ukr}_j$, valence and volume of slant about the Ukraine-crisis news, $V^-_j$ and $V^+_j$, and quality measured as $\bar{\alpha}_j$. To make the regression coefficients comparable, we normalize the standard deviation of the absolute value differences to 1. Table 12 presents the regression results. First, we can confirm that the ideological distance between the news outlets indeed has an effect on the correlations in the persistent preferences of the news outlets. For example, the distance between the news outlets in the valence of slant in the Ukraine-crisis coverage explains the most of the variation in the correlations between the news outlets, with 1 standard deviation more similar news outlets tending to have 5.56% more correlated persistent preferences of consumers. However, the
Figure 12: Posterior estimates of the correlation matrix of persistent consumer preferences for news websites, $\alpha_{ij} - \bar{\alpha}_i$.

Each dot represents the correlation of $\alpha_j - \bar{\alpha}_j$ for two news outlets. The scale on the right explains the color code of the correlations. The colors of the text labels correspond to types of the news outlets used throughout the draft (first explained in Figure 9).
quality of the news outlets also plays a role, with news outlets that are 1 standard deviation more similar in $\alpha_j$, having 4.22% more correlated persistent preferences of consumers.

Table 12: Relationship between the correlations in persistent preferences of consumers, $\alpha_{ij} - \bar{\alpha}_i$, and distance between the outlets’ characteristics.

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: cor($\alpha_{ij} - \bar{\alpha}<em>i, \alpha</em>{ij'} - \bar{\alpha}_{ij'}) \forall j \neq j'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.1155</td>
</tr>
<tr>
<td></td>
<td>(0.0191)</td>
</tr>
<tr>
<td>$</td>
<td>F_j^{IS} - F_{j'}^{IS}</td>
</tr>
<tr>
<td></td>
<td>(0.0076)</td>
</tr>
<tr>
<td>$</td>
<td>F_j^{Ukr} - F_{j'}^{Ukr}</td>
</tr>
<tr>
<td></td>
<td>(0.0076)</td>
</tr>
<tr>
<td>$</td>
<td>V_j^- - V_{j'}^-</td>
</tr>
<tr>
<td></td>
<td>(0.0085)</td>
</tr>
<tr>
<td>$</td>
<td>V_j^+ - V_{j'}^+</td>
</tr>
<tr>
<td></td>
<td>(0.0081)</td>
</tr>
<tr>
<td>$</td>
<td>\bar{\alpha}<em>j - \bar{\alpha}</em>{j'}</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
</tr>
<tr>
<td>Observations</td>
<td>630</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.2133</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.207</td>
</tr>
</tbody>
</table>

The results above suggest that persistent preferences of consumers, $\alpha_{ij}$, represent both the quality of the news outlet and some brand capital that the news outlet has accumulated over time, with an ideology of the news outlet playing a role in the brand capital formation.
10 Online Appendices

10.1 News Outlets Classification

News outlets' classification was done based on interviews with media professionals and the ownership structure of the news outlets. The ownership structure for January 2016 is presented below.

Government-controlled news outlets:

- vesti, 1tv, tass, rg, rt and ria are owned by the government.
- aif is owned by Moscow City Hall.
- ntv is owned by Gazprom, a state-owned gas monopolist.

- vz and dni were founded by Konstantin Rykov, a member of United Russia (incumbent political party) who led the political campaigns in support of Vladimir Putin in 2007. vz is owned by the Institute of Socio-Economics and Political Research, which is managed by Dmitry Badovsky, a former deputy chief of the Presidential Administration of Russia (2012).

Oligarchic news outlets:

- lenta and gazeta are owned by Alexander Mamut. Both were considered independent at the beginning of 2013. Gazeta changed its independent editor-in-chief to a more government-loyal editor-in-chief in September 2013; lenta underwent a similar change in March of 2014.67

- izvestia is owned by Yuri Kovalchuk through the National Media Group (NMG). Yuri Kovalchuk is a close friend of Vladimir Putin.

- lifenews is owned by Aram Gabrelyanov, a manager of NMG.68

- kommersant is owned by Alisher Usmanov, one of the richest Russian oligarchs.69

67 https://meduza.io/feature/2016/05/17/12-redaktsiy-za-pyat-let
68 http://www.kommersant.ru/doc/2311510
69 https://lenta.ru/lib/14164974/
- *kp* is owned by Grigory Berezkin, who is on the board of directors of state-owned RZD.\(^70\)

- *fontanka* is owned by “Azur-Media.”

Potentially government-influenced news outlets:

- *bfm* is owned by Rumedia, a company of Russian steel tycoon Vladimir Lisin.\(^71\)

- *echo* is jointly owned by journalists of echo (34%) and a state-owned gas monopolist Gazprom (66%). One of the most famous Russian independent media companies, it is reported to be influenced by the government and publish paid articles.\(^72\)

- *interfax’s* beneficiary is not disclosed, but there is information that it is owned by the top-management.\(^73\)

- *mk* is owned by Pavel Gusev, a confidant of Vladimir Putin. There are examples of mk removing published articles about government-sensitive topics.\(^74\)

- *znak* was formerly ura.ru; it had to change its name due to government pressure.\(^75\)

- *ng* is owned by Konstantin Remchukov. It is reported to publish articles which are paid for by the government.\(^76\)

- *polit’s, utro’s* and *ridus’s* ownerships are unclear.

- *regnum* is reported to have been purchased by Gazprom media.\(^77\) It is reported to publish paid articles.\(^78\)

- *rosbalt, sobesednik* and *trud* are reported to publish paid articles.\(^79\)

Independent news outlets:

\(^{70}\)http://www.forbes.ru/profile/grigorii-berezkin
\(^{71}\)https://en.wikipedia.org/wiki/Vladimir_Lisin
\(^{72}\)https://tjournal.ru/p/media-denim
\(^{73}\)https://www.vedomosti.ru/business/articles/2012/01/19/lgota_dlya_smi
\(^{74}\)http://www.rbc.ru/politics/27/12/2013/897386.shtml
\(^{76}\)http://theins.ru/politika/6015
\(^{77}\)https://lenta.ru/news/2014/06/20/media/
\(^{78}\)https://tjournal.ru/p/media-denim
\(^{79}\)https://tjournal.ru/p/media-denim
• newsru is owned by Vladimir Gusinsky, a tycoon who has opposed the incumbent Russian government since 2001.

• newtimes is owned by a non-profit fund The New Times Foundation.

• novayagazeta is owned by journalists (76%), Alexander Lebedev (14%) and Mikhail Gorbachev (10%).

• rbc and snob are owned by Mikhail Prokhorov, a Russian billionaire and politician. He run for president in the 2012 elections. RBC.ru stayed independent until May 2016, when the top managers were fired due to political pressure. It was later acquired by Grigory Berezkin, the owner of kp.ru, in June 2017.

• slon and tvrain are owned by Alexander Vinokurov and Natalia Sidneeva. tvrain’s TV channel was taken off the air by the major cable systems after covering 2011 street protests. Its website is subscription-based.

• vedomosti was jointly owned by Sanoma Independent Media (33%), Financial Times (33%) and The Wall Street Journal (33%) until the end of 2015. It was sold to Demyan Kudryavsev in November 2015 due to the a new law limiting foreign ownership of media to 20% starting in 2016.

• forbes was owned by Axel Springer before the end of 2015. It was sold to Alexander Fedotov in October 2015 due to a new law limiting foreign ownership of media to 20% starting in 2016.

• the-village is owned by Look at Media publishing.

International News Outlets:

• bbc is the Russian version of BBC.

• svoboda is Radio Liberty, a United States government-funded broadcasting organization.

• meduza is a news outlet founded in Latvia by a former journalists of lenta.ru, who were fired in March 2014 due to their Ukraine Crisis coverage.

• dw is the Russian version of Deutsche Welle.

---

• *reuters* is the Russian version of Reuters.

A small subset of Ukrainian news outlets: *korrespondent, liga* and *unian*.

### 10.2 Publication Records Collection and Processing

For the 48 outlets described in the Table 1, we collect information on their publications for the period starting April 1, 2013, and ending March 31, 2015. Data for the websites *fontanka.ru, izvestia.ru, ng.ru, svoboda.org, vedomosti.ru, slon.ru,* and *fontanka.ru* were collected from the media archive of *public.ru*. Data for the rest of the news outlets were scraped directly from the corresponding news websites. For the websites that did not provide an archive of the published articles, article URLs were collected from the media archive of *medialogia.ru*, and then these URLs were used to scrape the article information.

For all of the websites, information about the publication URLs, their dates and titles is available. For almost all of the websites, texts of the news publications are available, with 5 exceptions: *meduza.io, newtimes.ru, the-village.ru, snob.ru,* and *ridus.ru*. We use these websites only for the allocation of sensitive news and media slant in the news and exclude them from any other empirical exercises. When allocating the sensitive news, we treat titles of these 5 news outlets as texts of their articles.

To find sensitive news and the corresponding media slant, we process the texts of the news articles by stemming all the words and removing punctuation and stop words. We define proper nouns in the text corpus as any word that frequently (more than 50% of times used in the corpus) starts with a capital letter in the text when it is not at the beginning of the sentence.\(^\text{81}\)

### 10.3 Summary of Browsing Behavior

Each news website consists of 4 different types of pages: the main page, news articles pages, news subdirectories, and other pages. We classify the visit as the main page visit if the visited URL matches the main page url. We classify the visit as the news article visit if the visited URL matches one of the URLs of the publication records data or has a structure similar to it.\(^\text{82}\) We classify the URL as a subdirectory if the visited URL matches the subdirectory URL.

\(^\text{81}\)In doing this, we include the typical proper nouns but exclude words that are used as proper nouns rarely and only in a certain context.

\(^\text{82}\)For example, if the article URL has the structure http://www.x1.ru/news/topic/year/month/date/name-of-the-article.html, we classify any ULR with the structure http://www.x1.ru/news/topic/year/month/date/some-other-name-of-the-article.html as news articles.
We classify the rest of the URL visits as other page visits. The majority of the URL visits classified as other pages correspond to the photos, videos and other special content on news websites.

News articles account for most page views on news websites. Other webpages are visited half as often as news articles. The main directory and news subdirectories are also each visited only half as often as news articles. Table 13 shows statistics of browsing of the different webpage types. While some consumers read news from the headlines, most of the time the main pages and news subdirectories help readers to navigate to the news articles. This also includes navigation to the non-news content in the “other” sections. Thus, we only use navigation to news articles as records of news consumption.

Table 13: Summary of browsing behavior

<table>
<thead>
<tr>
<th>Webpage Type</th>
<th>Page views</th>
<th>Visits (Sessions)</th>
<th>Seconds spent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Main page</td>
<td>5,344,041</td>
<td>1917206</td>
<td>128</td>
</tr>
<tr>
<td>News articles</td>
<td>1,042,078</td>
<td>4240831</td>
<td>186</td>
</tr>
<tr>
<td>News subdirectories</td>
<td>4,225,221</td>
<td>1484410</td>
<td>263</td>
</tr>
<tr>
<td>Other</td>
<td>6,547,225</td>
<td>2,389,635</td>
<td>145</td>
</tr>
<tr>
<td>Total</td>
<td>26,537,267</td>
<td>6,630,400</td>
<td>176</td>
</tr>
</tbody>
</table>

For example, visits with a URL structure http://www.x1.ru/news/topic/.

---

83 For example, visits with a URL structure http://www.x1.ru/news/topic/.
10.4 Comparing Weekly Visitors of IE Toolbar and LI

Table 14 presents the visit shares of the 14 out of the top 30 websites in the scraped LI data. We exclude the seven news outlets described in the Table 4, news outlets that are split into multiple subsections in the LI data records, and news outlets that do not make the top 30 list more than half of the scraped days. For the resulting set of websites, we collect usage information for the news readers in the IE Toolbar data. IE Toolbar users are more likely to visit the weather predictions website, less likely to visit the entertainment websites such as movie descriptions and torrent trackers, more likely to visit odnoklassniki.ru, a social network popular with older audiences, and less likely to visit vkontakte.ru, a social network popular with younger audiences. This suggests that users of the IE Toolbar are older than the general population. It is also consistent with the notion that the IE Toolbar users are more likely to be office workers.

<table>
<thead>
<tr>
<th>Website</th>
<th>Description</th>
<th>Visit Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>auto.ru</td>
<td>Buy/Sell used cars</td>
<td>0.0141</td>
</tr>
<tr>
<td>avito.ru</td>
<td>Classified posts</td>
<td>0.0701</td>
</tr>
<tr>
<td>drom.ru</td>
<td>Website about cars</td>
<td>0.0170</td>
</tr>
<tr>
<td>gismeteo.ru</td>
<td>Weather</td>
<td>0.0347</td>
</tr>
<tr>
<td>hh.ru</td>
<td>Job postings</td>
<td>0.0138</td>
</tr>
<tr>
<td>kinopoisk.ru</td>
<td>Movie descriptions</td>
<td>0.0229</td>
</tr>
<tr>
<td>ngs.ru</td>
<td>Novosibirsk city website</td>
<td>0.0155</td>
</tr>
<tr>
<td>odnoklassniki.ru</td>
<td>Social Network (older audience)</td>
<td>0.2592</td>
</tr>
<tr>
<td>pluso.ru</td>
<td>Records clicks to social media</td>
<td>0.1186</td>
</tr>
<tr>
<td>rutracker.org</td>
<td>Torrent website</td>
<td>0.0182</td>
</tr>
<tr>
<td>tiu.ru</td>
<td>Online retailer</td>
<td>0.0140</td>
</tr>
<tr>
<td>vkontakte.ru</td>
<td>Social Network (younger audience)</td>
<td>0.3755</td>
</tr>
<tr>
<td>wildberries.ru</td>
<td>Online retailer</td>
<td>0.0137</td>
</tr>
<tr>
<td>woman.ru</td>
<td>Online magazine</td>
<td>0.0129</td>
</tr>
</tbody>
</table>

Unfortunately, we do not have information on the IE Toolbar users who are not the news readers. While our definition of the news readers is broad (visit a URL of the top 48 Russian online news outlets at least once over one and a half years), focusing only on browsing behavior of the news readers might lead to selection driving the differences between columns 3 and 4 of Table 14.
Figure 13: Normalized traffic of the top seven news websites, IE Toolbar and Liveinternet.ru

For each website and news source, the average traffic level is normalized to one, and the IE Toolbar data are corrected for the churn rate. The correlation between the traffic changes in the IE Toolbar and LI dataset is in the brackets.
10.5 Weekly Users of IE Toolbar

Figure 14: Normalized number of weekly visitors of IE Toolbar data
10.6 Sensitive News: Censorship and Slant

10.6.1 Censored unigrams and bigrams

Tables 15 and 16 present 54 bigrams of the proper nouns that are underused by the GC news outlets. To define a set of censored bigrams, we exclude the bigrams related to the profession of journalism, such as names of journalists, media owners, news outlets, etc. We also exclude three common actors, Dmitry Medvedev, Ramzan Kadyrov and Alisher Usmanov, given that there is a lot of regular news about these actors. The resulting set of censored bigrams of the proper nouns contain 34 bigrams (marked bold in the tables 15 and 16).

Table 15: List of the top 54 bigrams of the proper nouns underused by the GC news outlets. Part 1.

<table>
<thead>
<tr>
<th>Underused proper noun: English translation</th>
<th>Information about the proper nouns</th>
<th>Rank Difference, $\Delta \text{Rank}_{ind-Gov}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexei Navalny</td>
<td>Opposition politician</td>
<td>-28.3</td>
</tr>
<tr>
<td>(The) New Times</td>
<td>News outlet</td>
<td>-27.1</td>
</tr>
<tr>
<td>Mikhail Khodorkovsky</td>
<td>Opposition politician, political prisoner</td>
<td>-26.7</td>
</tr>
<tr>
<td>Echo (of) Moscow</td>
<td>News outlet</td>
<td>-26.6</td>
</tr>
<tr>
<td>Dmitry Kiselyov</td>
<td>Journalist</td>
<td>-26.6</td>
</tr>
<tr>
<td>Sergei Guriev</td>
<td>Economist, interrogated about “Yukos”</td>
<td>-25.8</td>
</tr>
<tr>
<td>Gennady Timchenko</td>
<td>Businessman, friend of Vladimir Putin</td>
<td>-25.7</td>
</tr>
<tr>
<td>Galina Timchenko</td>
<td>Journalist</td>
<td>-25.1</td>
</tr>
<tr>
<td>Svetlana Davydova</td>
<td>Civilian investigated for treason</td>
<td>-24.6</td>
</tr>
<tr>
<td>Alexander Plushev</td>
<td>Journalist</td>
<td>-24.4</td>
</tr>
<tr>
<td>Marat Gelman</td>
<td>Gallerist</td>
<td>-24.4</td>
</tr>
<tr>
<td>Alexei Navalny (2)</td>
<td>Opposition politician</td>
<td>-24.3</td>
</tr>
<tr>
<td>Ilya Yashin</td>
<td>Opposition politician</td>
<td>-24</td>
</tr>
<tr>
<td>Pussy Riot</td>
<td>Protest punk rock band</td>
<td>-23.2</td>
</tr>
<tr>
<td>Sergey Parkhomenko</td>
<td>Political journalist</td>
<td>-22.9</td>
</tr>
<tr>
<td>Alexei Venediktov</td>
<td>Editor-in-Chief of a News Outlet</td>
<td>-22.8</td>
</tr>
<tr>
<td>Alexander Vinokurov</td>
<td>Owner of multiple news outlets</td>
<td>-22.3</td>
</tr>
<tr>
<td>Arkady Rotenberg</td>
<td>Businessman, friend of Vladimir Putin</td>
<td>-22.3</td>
</tr>
<tr>
<td>Andrei Zubov</td>
<td>History professor</td>
<td>-22.2</td>
</tr>
<tr>
<td>Mikhail Kosenko</td>
<td>Political prisoner, Bolotnaya protests</td>
<td>-22.1</td>
</tr>
<tr>
<td>Alexei Kudrin</td>
<td>Politician, former minister</td>
<td>-21.9</td>
</tr>
<tr>
<td>The New (Times)</td>
<td>News outlet</td>
<td>-21.8</td>
</tr>
<tr>
<td>Igor Sechin</td>
<td>Chairman of Rosneft, close ally of Putin</td>
<td>-21.8</td>
</tr>
<tr>
<td>Ramzan Kadyrov</td>
<td>Head of the Chechen Republic</td>
<td>-21.5</td>
</tr>
<tr>
<td>(The) Other Russia</td>
<td>Opposition political party</td>
<td>-21.4</td>
</tr>
</tbody>
</table>

Bigrams marked as bold are selected to define sensitive news.
Table 16: List of the top 54 bigrams of the proper nouns underused by the GC news outlets.

Part 2.

<table>
<thead>
<tr>
<th>Underused proper noun:</th>
<th>Information about the proper nouns</th>
<th>Rank Difference, $\Delta \text{Rank}^{\text{Ind-Gov}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pavel Durov</td>
<td>Entrepreneur</td>
<td>-21</td>
</tr>
<tr>
<td>Cosmopolitan, Esquire</td>
<td>News outlets</td>
<td>-21</td>
</tr>
<tr>
<td>Echo Petersburg</td>
<td>News outlet</td>
<td>-21</td>
</tr>
<tr>
<td>Alexei Venediktov</td>
<td>Editor-in-Chief of a news outlet</td>
<td>-20.9</td>
</tr>
<tr>
<td>Yukos Capital</td>
<td>Former company of Michail Khodorkosky</td>
<td>-20.9</td>
</tr>
<tr>
<td>Alexei Navalny</td>
<td>Opposition politician</td>
<td>-20.9</td>
</tr>
<tr>
<td>The Village</td>
<td>News Outlet</td>
<td>-20.9</td>
</tr>
<tr>
<td>Kakha Bendukidze</td>
<td>Georgian politician</td>
<td>-20.9</td>
</tr>
<tr>
<td>Natalia Sidneeva</td>
<td>Editor of a news outlet</td>
<td>-20.7</td>
</tr>
<tr>
<td>Yves Rocher</td>
<td>Company from Alexey Navalny’s court case</td>
<td>-20.6</td>
</tr>
<tr>
<td>Nikolai Lyaskin</td>
<td>Manager of FBK, Alexei Navalny’s fund</td>
<td>-20.6</td>
</tr>
<tr>
<td>Anton Nosik</td>
<td>Media manager</td>
<td>-20.6</td>
</tr>
<tr>
<td>Svetlana Davydova</td>
<td>Civilian investigated for treason</td>
<td>-20.6</td>
</tr>
<tr>
<td>Irina Prohorova</td>
<td>Head of the opposition political party</td>
<td>-20.5</td>
</tr>
<tr>
<td>Mikhail Demin</td>
<td>Media Manager</td>
<td>-20.5</td>
</tr>
<tr>
<td>Yuri Saprikin</td>
<td>Journalist</td>
<td>-20.4</td>
</tr>
<tr>
<td>Alisher Usmanov</td>
<td>Billionaire</td>
<td>-20.4</td>
</tr>
<tr>
<td>Yulia Navalaya</td>
<td>Wife of Alexey Navalny</td>
<td>-20.2</td>
</tr>
<tr>
<td>Sergey Aleksashenko</td>
<td>Russian Economist</td>
<td>-20.2</td>
</tr>
<tr>
<td>Pavel Chikov</td>
<td>Head of the Human Rights Group Agora</td>
<td>-19.8</td>
</tr>
<tr>
<td>Platon Lebedev</td>
<td>Associate of Mikhail Khodorkovsky</td>
<td>-19.8</td>
</tr>
<tr>
<td>Denis Sinyakov</td>
<td>Photographer and political activist</td>
<td>-19.8</td>
</tr>
<tr>
<td>Yaroslav Belousov</td>
<td>Political prisoner</td>
<td>-19.2</td>
</tr>
<tr>
<td>Transparency International</td>
<td>International NGO</td>
<td>-19.2</td>
</tr>
<tr>
<td>Kira Yarmish</td>
<td>Press-secretary of Alexey Navalny</td>
<td>-19.1</td>
</tr>
<tr>
<td>Dmitry Medvedev</td>
<td>Prime Minister of Russia</td>
<td>-18.9</td>
</tr>
<tr>
<td>Lubov Sobol</td>
<td>Manager of FBK, Alexei Navalny’s fund</td>
<td>-18.9</td>
</tr>
<tr>
<td>Mikhail Lesin</td>
<td>Media manager</td>
<td>-18.9</td>
</tr>
<tr>
<td>Alexei Grazdankin</td>
<td>Deputy director of Levada Center</td>
<td>-18.8</td>
</tr>
</tbody>
</table>

Bigrams marked as bold are selected to define sensitive news.
In addition to the bigrams of the proper nouns, we re-do the classification using the unigrams of the proper nouns. We do this to make sure that we do not exclude facts described with a single proper noun. Figure 15 presents the histograms of the rank difference distributions, $h_{\text{ind-gov}}^{\text{actual}}$ and $h_{\text{ind-gov}}^{\text{random}}$. To define censored proper nouns we compare the lowest rank difference in $h_{\text{ind-gov}}^{\text{actual}}$ (-29.3) and in $h_{\text{ind-gov}}^{\text{random}}$ (-21.1). There are 47 unigrams of the proper nouns in the actual sample with a rank difference below the threshold of -21.1. A lot of these unigrams correspond to the last names of the sensitive actors which are classified based on bigrams, and some others refer to the ambiguous actors.

Figure 15: Histograms of $\Delta \text{Rank}_{\text{ind-Gov}}$ across the proper nouns: actual and random corpus.

![Histogram in the blue color corresponds to the actual corpus, histogram in the green color – to the random corpus. The red vertical line is a cutoff corresponding to the lowest rank difference in the random sample, -21.1.](image)

Table 17 provides an example of the top 20 underused unigrams. To define a set of censored unigrams, we exclude the unigrams related to the profession of journalism, and unigrams that refer to ambiguous actors. The resulting set of censored unigrams contains 10 proper nouns (marked bold in the table 17).
Table 17: List of the top 20 unigrams of the proper nouns underused by the GC news outlets.

<table>
<thead>
<tr>
<th>Underused proper noun</th>
<th>Information about the proper nouns</th>
<th>( \Delta \text{Rank}_{Ind-Gov} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venediktov</td>
<td></td>
<td>-29.3</td>
</tr>
<tr>
<td>Rotenberg</td>
<td></td>
<td>-29</td>
</tr>
<tr>
<td>Timchenko</td>
<td></td>
<td>-28.2</td>
</tr>
<tr>
<td>Slon</td>
<td>News outlet</td>
<td>-28.1</td>
</tr>
<tr>
<td>Revzin</td>
<td>Journalist</td>
<td>-27.9</td>
</tr>
<tr>
<td>Roskomnadzor</td>
<td>Federal agency overseeing media</td>
<td>-27.5</td>
</tr>
<tr>
<td>Khodorkovsky</td>
<td></td>
<td>-27.4</td>
</tr>
<tr>
<td>Venediktov</td>
<td></td>
<td>-27.2</td>
</tr>
<tr>
<td>Navalny</td>
<td></td>
<td>-26.4</td>
</tr>
<tr>
<td>Plushev</td>
<td></td>
<td>-25.7</td>
</tr>
<tr>
<td>Ketchum</td>
<td>PR agency of Russian government</td>
<td>-25.7</td>
</tr>
<tr>
<td>Echo</td>
<td></td>
<td>-25.6</td>
</tr>
<tr>
<td>Lebedev</td>
<td></td>
<td>-25.5</td>
</tr>
<tr>
<td>Kudrin</td>
<td></td>
<td>-25.1</td>
</tr>
<tr>
<td>Sechin</td>
<td></td>
<td>-24.9</td>
</tr>
<tr>
<td>Kosenko</td>
<td></td>
<td>-24.3</td>
</tr>
<tr>
<td>Bolotnaya</td>
<td>Square where protests take place</td>
<td>-24.3</td>
</tr>
<tr>
<td>Prohorov</td>
<td></td>
<td>-24.3</td>
</tr>
<tr>
<td>Shlosberg</td>
<td>Opposition Politician</td>
<td>-24.2</td>
</tr>
<tr>
<td>Sakharov</td>
<td>Ambiguous, might be multiple actors</td>
<td>-24.2</td>
</tr>
<tr>
<td>Bukovsky</td>
<td>Ambiguous, might be multiple actors</td>
<td>-23.9</td>
</tr>
<tr>
<td>Gelman</td>
<td></td>
<td>-23.8</td>
</tr>
</tbody>
</table>

Unigrams marked as bold are selected to define sensitive news.
### 10.7 Media Slant in the Ukraine Crisis News

Table 18: List of the words corresponding to the pro-Russia and pro-Ukraine slant.

<table>
<thead>
<tr>
<th>Overused words by the GC news outlets</th>
<th>Overused words Ukrainian news outlets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>$\Delta \text{Rank}_{\text{Ukr-Gov}}$</td>
</tr>
<tr>
<td>reunion</td>
<td>-34.67</td>
</tr>
<tr>
<td>radical</td>
<td>-34.10</td>
</tr>
<tr>
<td>punitive</td>
<td>-33.47</td>
</tr>
<tr>
<td>overturn</td>
<td>-33.07</td>
</tr>
<tr>
<td>blockade</td>
<td>-32.60</td>
</tr>
<tr>
<td>bombing</td>
<td>-32.20</td>
</tr>
<tr>
<td>coup</td>
<td>-31.73</td>
</tr>
<tr>
<td>anti-Russian</td>
<td>-31.10</td>
</tr>
<tr>
<td>bombing (2)</td>
<td>-30.80</td>
</tr>
<tr>
<td>russophobe</td>
<td>-30.57</td>
</tr>
<tr>
<td>ultra-nationalist</td>
<td>-30.53</td>
</tr>
<tr>
<td>neo-nazi</td>
<td>-30.47</td>
</tr>
<tr>
<td>intra-Ukrainian</td>
<td>-30.13</td>
</tr>
<tr>
<td>nazism</td>
<td>-30.03</td>
</tr>
<tr>
<td>russophobe (2)</td>
<td>-28.33</td>
</tr>
<tr>
<td>nazi</td>
<td>-27.50</td>
</tr>
<tr>
<td>reunion (2)</td>
<td>-27.33</td>
</tr>
<tr>
<td>neo-nazi (2)</td>
<td>-27.27</td>
</tr>
</tbody>
</table>
10.8 Ideological Positions of the News Outlets

Figure 16: Reporting about internally sensitive news, by news outlets’ types.

Each text string represents a position of a news outlet. We remove five news outlets for which we have only information about the titles and about the text of the articles.
Figure 17: Reporting about the Ukraine-crisis news by news outlet type.

Each text string represents a position of a news outlet. We remove five news outlets for which we have information only about the titles and about the text of the articles.

Figure 18: Ideological positions of the news outlets in the Ukraine-crisis news coverage.

Each text string represents an ideological position of a news outlet in the Ukraine crisis news coverage.

64
10.9 Share of News Censored and Volume of Sensitive News

To access whether censorship is more prominent on the days with more internally sensitive news, we regress the difference in reporting about internally sensitive news between the GC and independent news outlets, \( F_{\text{Ind}, t}^{IS} - F_{\text{Gov}, t}^{IS} \), on the share of articles about the internally sensitive news in the market, \( F_t^{IS} \). Subfigure (a) in Figure 19 shows that there is a positive correlation between \( F_{\text{Ind}, t}^{IS} - F_{\text{Gov}, t}^{IS} \) and \( F_t^{IS} \), indicating the GC news outlets indeed censor more articles on the days with a high volume of sensitive news. On average, the GC news outlets report around 30.3% of the amount of the internally sensitive news that the independent news outlets report, \( E\left( \frac{F_{\text{Gov}, t}^{IS}}{F_{\text{Ind}, t}^{IS}} \right) = 0.303 \). This ratio of reporting is constant in volume of the internally sensitive news, as is shown in Subfigure (b) of Figure 19. We conclude that the average reporting of the news outlets about the internally sensitive news, \( F_j^{IS} \), is a good proxy for their reporting about the internally sensitive news over time.

Figure 19: Difference in the news reporting of GC and independent outlets is bigger on days with more internally sensitive news, and the ratio is stable over time.

(a) Difference

(b) Ratio

The red line corresponds to the fitted values of the linear regression. Subfigure (a) corresponds to the linear regression of \( F_{\text{Ind}, t}^{IS} - F_{\text{Gov}, t}^{IS} \) on \( F_t^{IS} \); the slope coefficient is statistically significant \( (p < .001) \). Subfigure (b) corresponds to the linear regression of \( \frac{F_{\text{Gov}, t}^{IS}}{F_{\text{Ind}, t}^{IS}} \) on \( F_t^{IS} \); the slope coefficient is not statistically significant \( (p = .221) \).
10.10 MCMC Estimation

We estimate the demand parameters by simulating from the posterior distribution defined in Section 6.2 and use the choice data from a set of consumers defined in Section 6.1. Given the large amount of choices that consumers make and a substantial number of choice alternatives (37), the estimation process requires substantial time and RAM memory resources. We use the hybrid MCMC sampler \texttt{rhierMnlRwMixture} from the 3.1 version of the \texttt{bayesm} package in R (Rossi et al., 2005). The sampler in this package is written in Rcpp, an R library that allows the integration of R and C++ languages, which substantially speeds up the computational process. Still, given the number of consumer choices and alternatives, estimation takes a significant amount of time. We run the MCMC procedure for the full sample of consumers for 20,000 iterations, storing every twenty-fifth observation for the memory reasons, leading to 800 saved MCMC draws. Estimation is complete in 12 days and requires 800 Gb of RAM. Figure 20 shows the log likelihood of the MCMC draws. We throw away the first 100 saved draws (2,500 actual draws) to remove the effect of the starting point on the sampler. Figure 21 shows an overview of the evolution of the $E(\theta)$ draws.\textsuperscript{85} We treat the last 700 MCMC draws as draws from the stationary distribution.

Figure 20: Log likelihood of the 800 kept MCMC draws.

\begin{center}
\includegraphics[width=0.5\textwidth]{figure20.png}
\end{center}

The red line corresponds to the first 100 MCMC draws that are discarded.

\textsuperscript{85}We omit the draws of $\eta_i^*$ since it simply ensures that consumers never revisit the news outlet on the same day.
Figure 21: MCMC draws of $E(\theta)$.