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

Andrey Simonov
Columbia University

Justin Rao
HomeAway, Inc.

9th August 2017
First draft: 25th January 2016

Anecdotal evidence suggests consumers navigate to news outlets that are government controlled (GC) and pro-government biased even in the presence of independent news outlets. Does this imply a demand for pro-government bias in the news? Or do consumers enjoy other aspects of GC news outlets such as quality or brand? To answer these questions, we examine consumers' demand in the Russian online news market. We use publication records of the top 48 online news outlets to characterize the methods of government control in the news and identify the news that is sensitive for the government. We then use temporal variation in the amount of sensitive news and click-level browsing panel data to estimate the demand for news. The average consumer prefers the quality and brand of the GC news outlets but the news coverage with less pro-government bias, and there is a substantial heterogeneity in preferences. In a counterfactual simulation, we show that the GC news outlets would have 16.9% higher market shares in the absence of control and 42.9% lower market shares if their quality was at the level of the independent news outlets. Higher quality of the GC news outlets expands their share of online attention from 21.98% to 33.1%, substantially increasing their media power.

1   Introduction

On August 23, 2016, BBC.com published a news story covering a 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 discussed that an arbitration court upheld an earlier decision to exclude 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 using doping.

Such 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 2014). In case of coverage of the Russian Paralympic team ban, there is strong evidence suggesting that media slant comes from the supply side: website RT.com is owned by the Russian government, and "RT" is short for "Russia Today". However, this government-induced bias is different from the government capture in the model of Besley and Prat (2006), where a government needs to capture all news sources to suspend information dissemination. If other news outlets cover the topic truthfully, and consumers have access to all the news outlets, why would the Russian government invest money in RT? Does it attract (and potentially persuade) customers with different ideological views by its quality or brand, or is it read only by the customers whose political beliefs are similar to the political position of the Russian government?

In this paper, we aim to shed light on the above questions using the Russian online news market in 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 the news outlets being easily accessible to any user of the internet. A stylized fact in this market is that the GC news outlets enjoy high and stable market shares in this period of time: around 25% of news outlets are GC (owned by the government), and their overall market share is around 35-40%.[3] The key question of this paper is what drives demand for the GC news outlets

---

[1] http://www.bbc.com/sport/disability-sport/37165427

[2] https://www.rt.com/sport/356863-paralympic-russia-reaction-rio/

[3] Based on the statistics from http://www.liveinternet.ru. Historical data is scrapped using the Way-

in Russia. To formalize the motivating example of RT above, we distinguish between two potential families of explanations. On the one hand, 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, or because they value knowing the ideological position of the government, or because their find ideologically-heavy news content entertaining. On the other hand, consumers might have a distaste for the pro-government ideology in the news but have a persistent preference for the GC news outlets, reflecting higher quality or a superior brand of the GC news outlets.

To distinguish between these alternative explanations for the demand for GC news outlets in Russia, we collect two novel datasets. First, we scrape all the publication records for the top 48 online news outlets that write 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 find 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"), president Putin's affiliates related to corruption (e.g., "Rotenberg," "Timchenko"), and political protests (e.g. "Bolotnaya"). We label censored news as "internal-sensitive news", and characterize the ideological positions of the news outlets as a share of internal-sensitive news that they report. Second, we examine news coverage of the Ukraine crisis of 2013-2015, a sensitive news topic widely reported to have a pro-Russia propaganda. We show that the GC news outlets systematically report more news about the Ukraine crisis compared to the independent news outlets, but these news exhibit the pro-Russia (or 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 east-

---

back Machine: `http://web.archive.org/`

2

ern Ukraine. Similarly, the Ukrainian news outlets with coverage in Russian systematically overuse pro-Ukraine (or anti-Russia) slant: for example, 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 as a share of slanted Ukraine-crisis news articles.

Having the ideological positions of the news outlets, we build and estimate the demand model for news. Consumers have persistent preferences for the news outlets, preferences for the news topics and preferences for the ideological slant in the 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 the 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. Thus, our identification strategy relies on the exogenous changes in the number of sensitive news events happening over time. On days with no sensitive news, censorship and propaganda constraints do not affect the GC news outlets' coverage, making it similar to the coverage of the independent news outlets. On such days consumption is driven 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 the 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 are more likely to visit both the GC and independent news outlets on days with sensitive news, consumers behave like they value knowing the government's position about sensitive issues.[4] Finally, if consumers start navigating to more slanted news outlets on days with a lot of sensitive news, regardless of the valence of this bias, consumers are simply entertained by the slanted news.

Estimates of the structural model reveal a nuanced picture. On the one hand, an average consumer prefers the overall ideological position of the independent news outlets and 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

---

[4]Mullainathan and Shleifer (2005) refer to these consumers as "conscientious" news readers.

the GC news outlets. In particular, 37% of consumers prefer the ideological slant of the GC news outlets in coverage of the Ukraine crisis, and 50.3% of consumers prefer the coverage of the GC news outlets about the internal-sensitive topic. To test whether the GC news outlets are worse off under their idealogical positions, we perform several counterfactual simulations, changing the ideological positions of the news outlets and their quality level. In the absence of direct control, the GC news outlets would get 17% higher market shares, corresponding to a rough back-of-the-envelope estimate of $15.7 million in advertising revenues. In contrast, if the average quality of the GC news outlets was similar to the independent news outlets, they would get 42.8% lower market share ($39.5 million). Thus, we conclude that quality is the main reason for the observed demand for the GC news outlets. We extend our counterfactual simulations to examine the importance of the indirect government control in this market (Gehlbach and Sonin 2014) and show that it also has a limited effect on the market shares of the government-influenced companies.

In light of the importance of quality or brand for the demand for the GC news outlets in this market, we discuss the implied media power (Prat, 2017) induced by this quality of the GC news outlets. We show that the higher quality of the GC news outlets increases their share of online attention from 21.98% to 33.1%, 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 19.7%-29.7% 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 the alternative explanations for the nature of the demand for bias. We find that the majority of consumers prefer news about the Ukraine crisis with lower pro-government valence (60.42%) and lower volume of slant (51.78%). The vast majority of consumers, 72.24%, prefer more ideologically-similar news sources on days with more Ukraine-crisis news. Thus, only a minority (27.76%) 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 method to measure media slant (Groseclose and Milyo 2005, Gentzkow and Shapiro 2010, Gentzkow et al. 2016), which to our knowledge is the first measure that splits media slant into multiple

---

[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.

dimensions.[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 2015) and online news markets (Gentzkow and Shapiro 2011, Gentzkow and Shapiro 2015, Sen and Yildirim 2016, Athey et al. 2017, Cage 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 2014). Finally, our work is related to the theoretical (Besley and Prat 2006, Prat and Strömberg 2013, Gehlbach and Sonin 2014) and empirical (Durante and Knight 2012, Enikolopov et al. 2011, Bai et al. 2014, Roberts 2014, Garcia-Arenas 2016, Knight and Tribin 2016) literature on the effect of government control of the news on consumers, and 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 estimation procedure and the results. We run the counterfactual simulations of changing the level of government control in Section 7. Section 8 concludes.

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 the 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 change their chief editors.[7] The most prominent examples include dissolution of RIA Novosti, a state news agency known for balanced news coverage under its editor-in-chief Svetlana Mironyuk, in

---

[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 Chinese ideological spectrum is multi-dimensional.

[7]ura.ru - November 30, 2012, change of chief editor, `http://www.vesti.ru/doc.html?id=972487` (Russian); gazeta.ru - September 7, 2013, change of chief editor, `http://slon.ru/russia/grekh_unyniya_kreaklov_gazete_ru_naznachili_novyy_glavnyy_redaktor-988192.xhtml` (Russian); ria.ru - December 10, 2013, liquidation, chief editor fired, `http://www.rg.ru/2013/12/09/ykaz-dok.html` (Russian); lenta.ru - March 12, 2014, change of chief editor, `http://lenta.ru/news/2014/03/12/goreslavsky/` (Russian)

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 developing a law to limit the foreign ownership of Russian news outlets to 20% starting in January 2016.[11]

Table 1: Russian-language online news media by the type of influence in December 2014

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

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 open information about ownership structure and interviews with media professionals.

In 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 popular Russian-language sections. Russian news outlets are organized into four groups: (1) independent and not influenced, (2)

---

[8]http://www.telegraph.co.uk/news/worldnews/europe/russia/10505386/Vladimir-Putin-dissolves-Russias-RIA-Novosti.html

[9]http://www.bbc.com/news/world-europe-26543464

[10]http://www.bbc.com/news/technology-26578264

[11]http://www.rg.ru/2014/10/17/smi.html (Russian); http://www.squirepattonboggs.com/~/media/files/insights/publications/2014/10/russia-moves-to-limit-foreign-ownership-in-the-2_ _/files/russiamovestolimitforeignownership/fileattachment/russiamovestolimitforeignownership.pdf

[12]Top outlets are defined using their market share on liveinternet.ru.

independent but possibly influenced, (3) possibly influenced news media owned by oligarchs close to 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 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 (2 news outlets). Appendix 9.1 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 face any 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 formally 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 news agenda directly controlled by the Kremlin. The majority of these news outlets are owned by the government and receive government subsidies.

2.2 Supply Data

For 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 *medialogia.ru* and *public.ru*. The resulting panel contains 3.9 million publications. For each article, we collect the title, text,[13] URL link, and timestamp. Table 2 presents the number of articles per type of news outlet. Appendix 9.2 provides more information on the publication records data collection and processing.

2.3 Demand Data

To measure news consumption, we use the Internet Explorer (IE) Toolbar browsing data, which include 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.[14] IE

---

[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.

[14]Based on Microsoft records, around 75% of users who installed the plug-in opt-in to the data collection.

Table 2: Number of articles by type of news outlet

Type	Articles
GC	1,168,569
Independent	494,087
Influenced	1,848,556
International	75,596
Ukrainian	315,927
Total	3,902,735

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.

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 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[15] 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 whose 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 IE Toolbar data, we classify the webpages that consumers visit in four groups: main pages of the websites, news subdirectories, news articles, and other pages (such as special projects, photos, videos, etc.).[16] Table 12 shows summary statistics of browsing by types of the 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%.

---

[15]Data for the period between April 1, 2013, and November 15, 2013, are available with scrubbed (deleted) user IDs.

[16]Appendix 9.3 contains details on classification.

Table 3: Summary of browsing behavior

	Page views	% of Page Views	Seconds spent	
			Mean	Median
Main page	5,344,041	20.1%	128	42
News articles	10,420,780	39.3%	186	86
News subdirectories	4,225,221	15.9%	263	90
Other	6,584,713	24.7%	145	44
Total	26,537,267	100%	176	64

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 of time studied for only the top 30 websites on the Russian internet, which includes seven news websites from our sample.[17]

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, and the product positioning of rbc.ru as a news agency focusing on the business news. This would explain both a higher visit shares of the rbc.ru and a lower visit share of the ria.ru, since ria.ru is a rival news agency in a direct competition with rbc.ru. Table 13 in the Appendix 9.4 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.[18]

---

[17]The top page includes only the top 30 websites; Wayback Machine does not have frequent records for the other pages.

[18]Please see the Appendix 9.4 for a more detailed comparison.

Table 4: Comparison of website rankings in IE Toolbar and LI.ru

Website	liveinternet.ru		IE Toolbar	
	Visit Share	Ranking	Visit Share	Ranking
rbc.ru	0.1951	1	0.3165	1
ria.ru	0.1800	2	0.0570	14
vesti.ru	0.1550	3	0.1879	2
kp.ru	0.1355	4	0.1146	4
lenta.ru	0.1319	5	0.1094	8
gazeta.ru	0.1248	6	0.1010	5
rg.ru	0.1240	7	0.1135	3

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 LI and IE Toolbar data.[19] Changes in the news consumption in IE Toolbar data closely track the population-level consumption in the LI data, with the correlation of 0.858. Figure 13 in the Appendix 9.4 present 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 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).

# 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 from the news coverage of its controlled outlets, or allows reports of only certain facts. 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

---

[19]For each website and news source, the average traffic level is normalized to one, and IE Toolbar data are corrected for the churn rate. The traffic is then averaged across the news outlets. Figure 14 in the Appendix 9.5 contains information about the number of weekly users of the IE Toolbar.

Figure 1: Normalized number of weekly visitors to rbc.ru, IE Toolbar and LI data
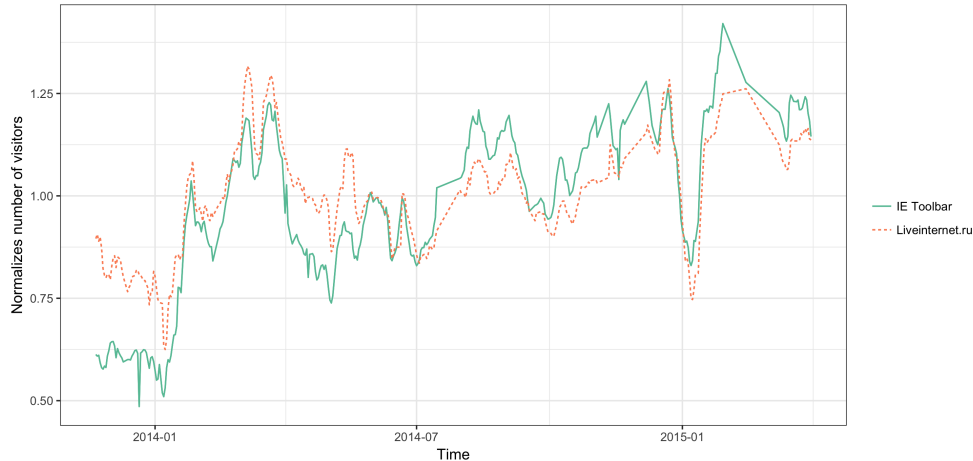
For each website and news source, the average traffic level is normalized to one, and IE Toolbar data are corrected for the churn rate. The traffic is then averaged across the news outlets.

classifies 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. 2015). 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).[20] In the media economics literature, propaganda corresponds to "framing and idealogical stand bias" (Prat and Strömberg 2013) and "distortion of news" (Genzkow et al. 2015).

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 news outlet's quality, representing all persistent characteristics of the news outlet. Government censorship affects which topics controlled news outlets report, and government propaganda affects the degree of pro-government slant in the sensitive news topics reporting.

---

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

3.2 Internal-Sensitive News

We use the definitions of government control and publication records data to find sensitive news topics. For this, we treat news articles as collections of words, simplified with standard natural language processing techniques such as stemming and removal of stop words.[21]

We identify the first set of government-sensitive news using the idea of censorship. Given that censorship is omission of facts, we focus on proper nouns, words that are likely to correspond to facts in the news.[22] For example, proper nouns represent the actors in the news and the places where the news happened.[23] We consider all words starting with 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 outlet compared to all the independent news outlet. 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 : $\text{sh}_{vj} = \frac{\text{count}_{vj}}{\sum_v \text{count}_{vj}} \ \forall \ v, j$, where count_{vj} is a number of occurrences of v in j's coverage

2. For each v, rank 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 $\text{sh}_{vj'} = max_{j \neq j'}(\text{sh}_{vj})$
   etc.

3. For each v, compute an average rank for the GC and independent news outlets:
   $\text{Rank}_v^x = \frac{\sum_{j \in x} \text{rank}_{vj}}{\sum_{j \in x} 1}$

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 $\Delta\text{Rank}_v^{Ind-Gov} = \text{Rank}_v^{\text{Ind}} - \text{Rank}_v^{\text{Gov}}$ are likely candidates for sensitive censored news topics. To test if low values of $\Delta\text{Rank}_v^{Ind-Gov}$ occur by chance, we randomly re-assign word counts within the news out-

---

[21]Appendix 9.2 provides more information on processing of the publication records data.

[22]Franceschelli (2011) also uses proper nouns to define the news topics. Cage et al. (2017) takes a more sophisticated topic detection algorithm.

[23]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," that summarize the topic of the news article, but does not capture the sentiment of this topic (captured by the word "worst").

lets and re-do the procedure.[24] Distribution of $\Delta\text{Rank}_v^{Ind-Gov}$ from the re-assigned scenario provides a benchmark distribution of the differences in ranks.[25] We define the empirical distribution of $\Delta\text{Rank}_v^{Ind-Gov}$ 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 13,514 bigrams of the proper nouns that appear more than 200 times in the news publications in our sample period.[26] Figure 2 presents the histograms of the rank difference distributions, $h_{\text{ind-gov}}^{\text{actual}}$ and $h_{\text{ind-gov}}^{\text{random}}$. Bigrams on the left side of the $h_{\text{ind-gov}}^{\text{actual}}$ distribution (negative rank difference) correspond to the potentially censored proper nouns. To test if these differences in ranks could occur by chance, we compare $h_{\text{ind-gov}}^{\text{actual}}$ to $h_{\text{ind-gov}}^{\text{random}}$. The lowest rank difference in $h_{\text{ind-gov}}^{\text{actual}}$ is -28.3, while in $h_{\text{ind-gov}}^{\text{random}}$ it is -18.8, corresponding to the red dashed line in the Figure 2.[27] 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.

We use the lowest rank difference in $h_{\text{ind-gov}}^{\text{random}}$, -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. To provide an example of these proper nouns, Table 5 presents 10 bigrams with the lowest $\Delta\text{Rank}_v^{Ind-Gov}$ difference. This list includes two prominent opposition politicians, an economist who had to flee Russia after a politically-motivated interrogation, a businessman in Putin's "Inner Circle" under the US sanctions, a mother of seven investigated for treason for making a call to the Ukrainian embassy, two news outlets and three journalists. Tables 14 and 15 in the Appendix 9.6 present all 54 censored bigrams. Excluding actors related to the profession of journalism (such as journalists, news outlets, media owners, etc.), all of these proper names are actors related to facts that are anecdotally known to be sensitive for the government.

We use the underused bigrams of the proper nouns to define the first group of sensitive

---

[24]For example, if a news outlet used only three words A, B, and C, and these words were used $\text{count}_A = 10$, $\text{count}_B = 15$ and $\text{count}_C = 20$ times, random re-assignment of word counts within a news outlet will permute the observed counts, for example $\text{count}'_A = 20$, $\text{count}'_B = 10$, $\text{count}'_C = 15$. 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.

[25]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.

[26]Threshold of 200 times is chosen arbitrarily.

[27]To make sure that the difference in the lowest $\Delta\text{Rank}_v^{Ind-Gov}$ 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 $h_{\text{ind-gov}}^{\text{random}}$.

Figure 2: Histograms of $\Delta\text{Rank}_v^{Ind-Gov}$ across bigrams of the proper nouns: actual and random corpus.
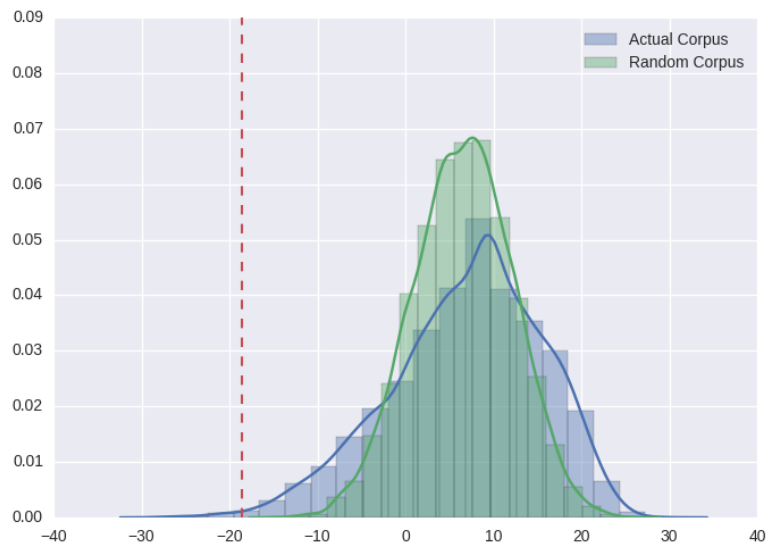
Histogram in blue color corresponds to the actual corpus, histogram in green color – to the random corpus. Red vertical line is a cutoff corresponding to the lowest rank difference in the random sample, -18.8.

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

Underused proper noun: English translation	Information about the proper nouns	Rank Difference, $\Delta\mathrm{Rank}_v^{Ind-Gov}$
Alexei Navalny	Opposition politician	-28.3
(The) New Times	News outlet	-27.1
Mikhail Khodorkovsky	Opposition politician, political prisoner	-26.7
Echo (of) Moscow	News outlet	-26.6
Dmitry Kiselyov	Journalist	-26.3
Sergei Guriev	Economist, interrogated about "Yukos"	-25.8
Gennady Timchenko	Businessman, friend of Vladimir Putin	-25.7
Galina Timchenko	Journalist	-25.1
Svetlana Davydova	Civilian investigated for treason	-24.6
Alexander Plushev	Journalist	-24.4

news. After excluding proper nouns related to the profession of journalism, we are left with 34 bigrams of the proper nouns which we define as censored. 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 extra 10 censored proper nouns based on that classification.[28] We call the news related to these censored bi- and unigrams of proper nouns "internal-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 contain at least one of the censored bi- and unigrams as an article about an internal-sensitive news topic.

3.2.1 Slant in the Internal-Sensitive News

We have found the internal-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. Apart from censorship, the GC news outlets might report internal-sensitive news with a slant different from the independent news outlet.

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

Figure 3 presents the histogram of $h_{\mathrm{ind\text{-}gov}}^{\mathrm{actual}}$ and $h_{\mathrm{ind\text{-}gov}}^{\mathrm{random}}$ for this corpus. The two distributions are almost identical. We use the cutoff mechanism described above to test if there

---

[28]See the Appendix 9.6 for more details.

is some systematic slant in the internal-sensitive news and find little evidence for this: out of 37,734 words in the corpus, only four word 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).[29] We conclude that we do not find evidence that the slant in the coverage of internal-sensitive news by the independent and GC news outlets is different.

Figure 3: Histograms of $\Delta\text{Rank}_v^{Ind-Gov}$ across words in the internal-sensitive news topics corpus.
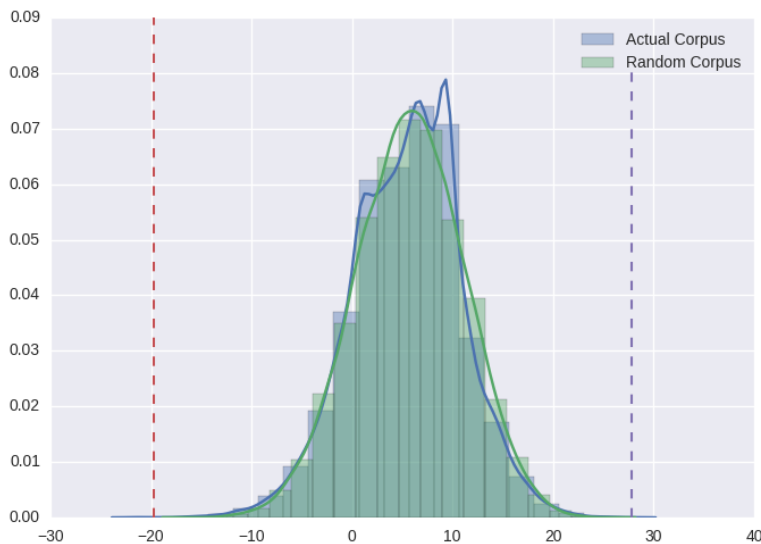
Histogram in blue color corresponds to the actual corpus, histogram in green color – to the random corpus. Red vertical line is a cutoff corresponding to the lowest rank difference in the random sample, and blue vertical line is a cutoff corresponding to the highest rank difference in the random sample.

3.3 Government-Sensitive News about the Ukraine Crisis

We now turn to another government-sensitive news topic: the Ukraine crisis of 2013-2015.[30] The conflict was widely covered in the Russian news media, and was reported to be heavily

---

[29]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 since they are related to the journalism itself and not to the news covered.

[30]For a broad overview, please see `https://en.wikipedia.org/wiki/Ukrainian_crisis`.

slanted by news outlets controlled by the Russian government.[31]

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" that were published in the independent, government-influenced, and GC outlets over time.[32] With the beginning of the Ukraine conflict, the figure shows that all news outlets increase their reporting about Ukraine, but GC outlets increase it more than independent and influenced outlets. This increase is almost the opposite of censorship: the GC news outlets report significantly more on the Ukraine crisis than the independent and influenced news outlets.

Figure 4: Share of articles containing the word "Ukraine" in the weekly coverage of news outlets, by types
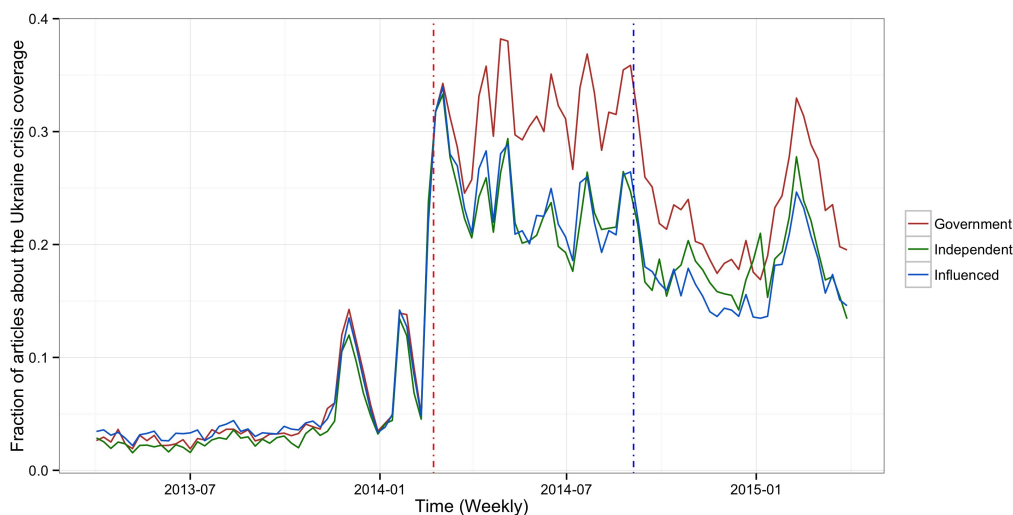
Red line corresponds to the GC media, green line - to the independent media, blue line - to the government-influenced media. Red dotted line corresponds to February 22, 2014, day when the former president Yanukovych fled Ukraine as a result of a revolution. Blue dotted line corresponds to the first Minsk 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 suffer from both pro-Russia and pro-Ukraine media slant.[33] Anecdotally, pro-Russian slant is framing

---

[31]For an overview, please see `https://en.wikipedia.org/wiki/Media_portrayal_of_the_Ukrainian_crisis#Media_in_Russia`.

[32]Having the word "Ukraine" in the news coverage is a proxy for an article being about the Ukraine crisis.

[33]For example, difference in the media slant is discussed on the fact-checking website `stopfake.org`,

the new Ukrainian government as a "fascist junta" that is conducting a "punitive operation" against the "rebels" in the Eastern Ukraine, and pro-Ukraine slant is framing Russia as an "aggressor" that has "occupied" the Ukrainian territory and supports "terrorist" and "separatists" in the Eastern Ukraine.

To test if there is 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 $\Delta\text{Rank}_v^{Ukr-Gov}$ across words in the Ukraine-crisis news topic corpus.
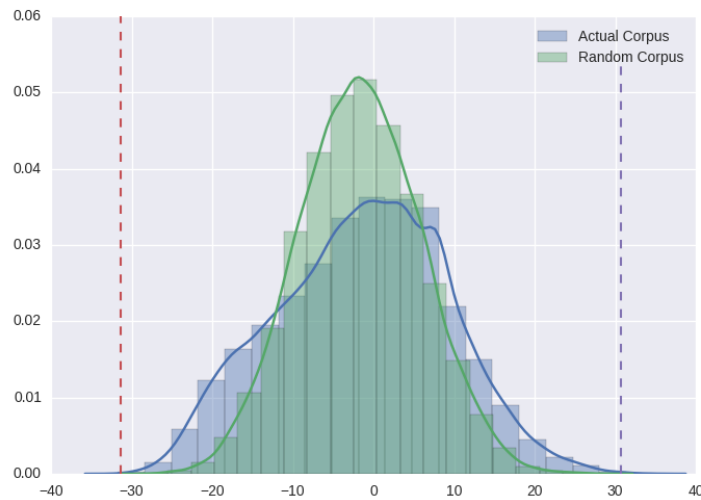
Histogram in blue color corresponds to the actual corpus, histogram in green color – to the random corpus. Red vertical line is a cutoff corresponding to the lowest rank difference in the random sample, and blue vertical line is a cutoff corresponding to the highest rank difference in the random sample.

18

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".[34] 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 Ukraine army is conducting an "anti-terrorist" operation against the "separatists".[35] We take this consistency as the evidence that the overused words indeed correspond to the pro-Russia and pro-Ukraine slant. Using the overused words and the anecdotal evidence, 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 17 in the Appendix 9.7 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.

Overused words by the:			
GC news outlets		Ukrainian news outlets	
Word	$\Delta\text{Rank}_v^{Ukr-Gov}$	Word	$\Delta\text{Rank}_v^{Ukr-Gov}$
reunion	34.7	continental	-31.7
radical	34.1	annexation	-30.8
punitive	33.5	monopolistic	-30
overturn	33.1	anti-terrorist	-29.9
blockade	32.6	devoid	-29.8
bombing	32.2	titushky[36]	-29.4
coup	31.7	content	-29.3
anti-Russian	31.1	residue	-29.3
colored	31	occupied	-29.3
deepest	31	deduced	-29.2

---

[34]The word "rebels" is the fifty-eights overused word with a rank difference of 27.3

[35]The word "separatists" is the twenty-first overused word with a rank difference of -28.4

[36]Ukrainian word to describe protesters supporting the former Ukraine government.

3.4 Reporting about the Sensitive News

3.4.1 Internal-Sensitive News

Knowing the identities of articles about internal-sensitive news and slant, we now characterize the reporting of news outlets. Figure 8 presents the average share of articles about the internal-sensitive topic by types of the news outlets. By construction, the independent news outlets report more about the internal-sensitive news than the GC outlets. The influenced news 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. Position of the international news outlets is closer to the independent news outlets, and Ukrainian news outlets tend to report very little about the internal-sensitive news, given that their coverage is focused on the issues in the Ukraine.

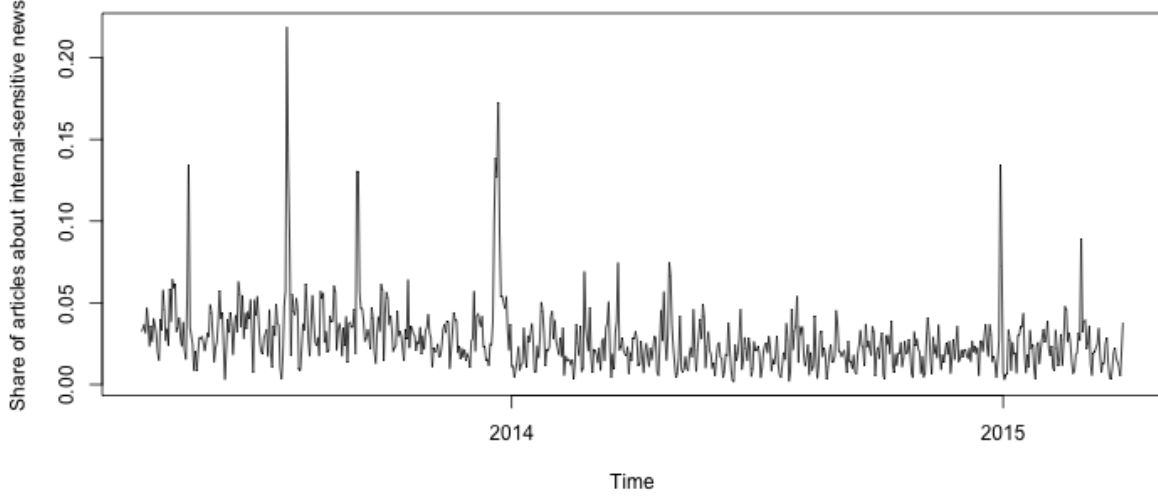Figure 6: Reporting about internal-sensitive news, by news outlets' types.
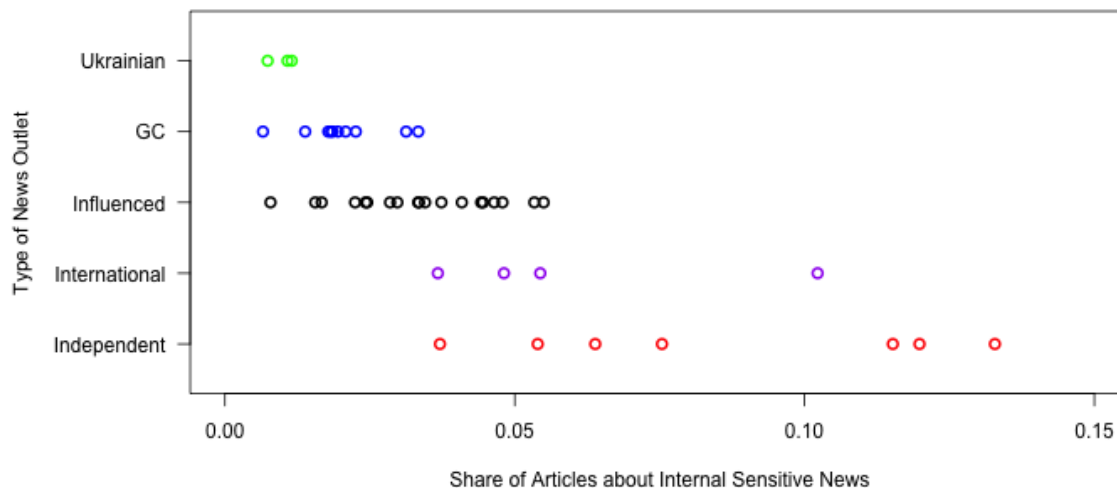
Each dot represents a position of a news outlet. We remove five news outlets for which we have only information about titles and about the text of the articles.

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 important on the days with more internal-sensitive news. To measure the volume of the internal-sensitive news, we compute the share of articles about the internal-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, there is 2.7% of articles about the internal-sensitive news. There is

also a substantial variation in the volume of internal-sensitive news across days, with F_t^{IS} varying from 0.2% to 21.9%.

Figure 7: Changes in the share of articles about the internal-sensitive news over time.

To access whether censorship is more important on the days with more internal sensitive news, we regress the difference in reporting about internal-sensitive news between the GC and independent news outlets, $F_{Ind,t}^{IS} - F_{Gov,t}^{IS}$, on the share of articles about the internal-sensitive news in the market, F_t^{IS}. Subfigure (a) in Figure 8 shows that there is a positive correlation between $F_{Ind,t}^{IS} - F_{Gov,t}^{IS}$ and F_t^{IS}, indicating the GC news outlets indeed censor more topics 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 internal-sensitive news that the independent news outlets report, $E(\frac{F_{Gov,t}^{IS}}{F_{Ind,t}^{IS}}) = 0.303$. In Subfigure (b) of Figure 8, we test whether the ratio of reporting on the internal-sensitive news, $\frac{F_{Gov,t}^{IS}}{F_{Ind,t}^{IS}}$, changes on the days with more sensitive news, and do not find significant correlation between the two. Thus, the average reporting of the news outlets about the internal-sensitive news, $\bar{F}_j^{IS}$, is a good proxy for their reporting about the internal-sensitive news over time.

3.4.2 Ukraine-crisis News

We now examine the reporting of news outlets about the Ukraine-crisis news. Figure 9 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

21

Figure 8: Difference in the news reporting of GC and independent outlets is bigger on days with more internal-sensitive news, and ratio is stable over time.
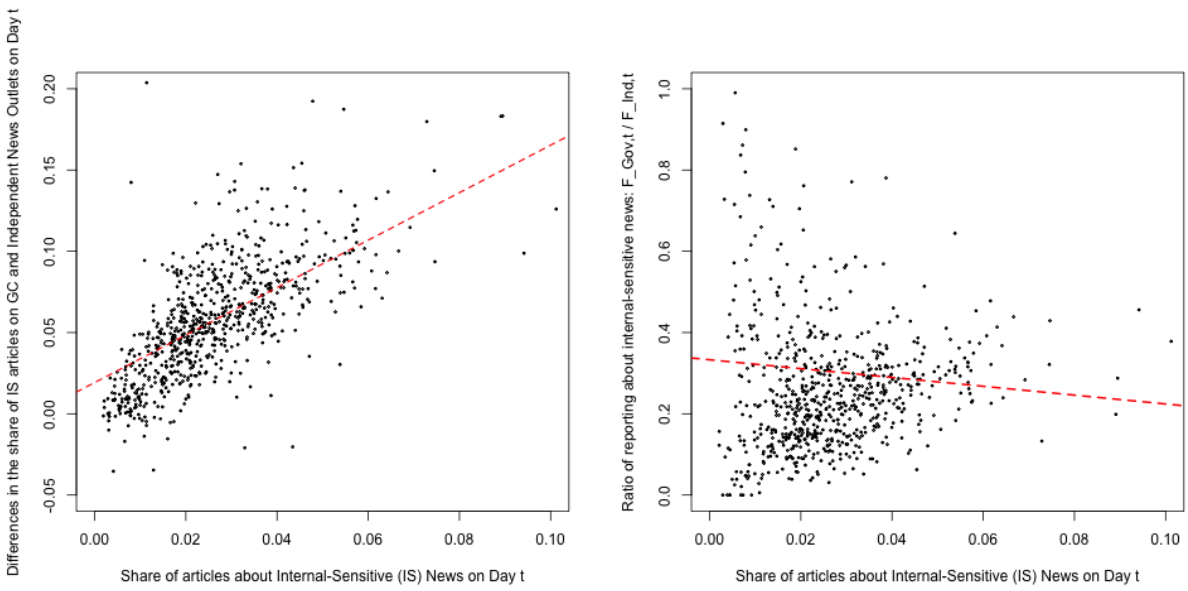
(a) Difference

(b) Ratio

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

news about the Ukraine crisis. International and Ukrainian news outlets report more about the Ukraine crisis.

Figure 9: Reporting about the Ukraine-crisis news, by news outlets' types.
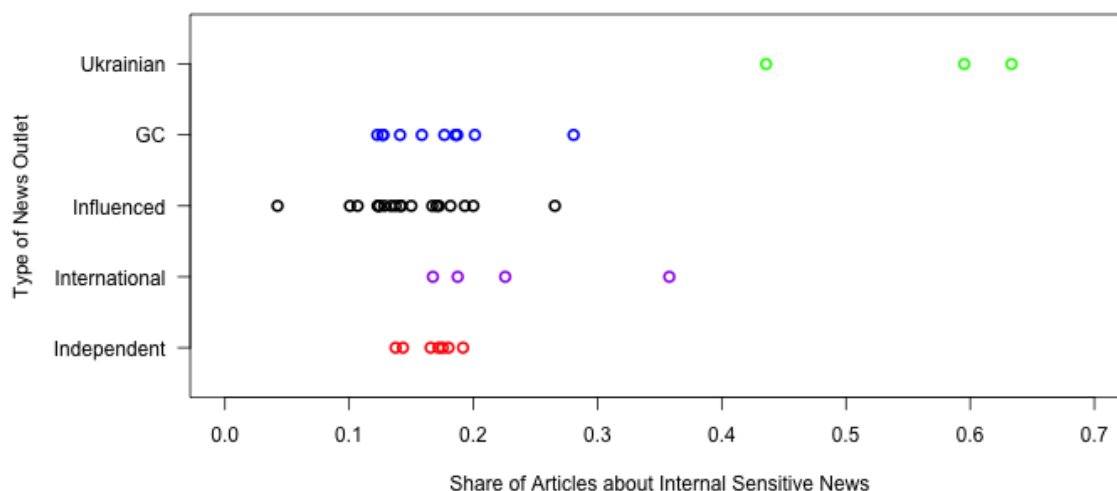
Each dot represents a position of a news outlet. We remove five news outlets for which we have only information about titles and about the text of the articles.

Subfigure (a) in Figure 10 presents the average shares of articles about the Ukraine crisis that have the pro-Russia or pro-Ukraine slant by the news outlets. By construction, the GC news outlets have relative high levels of the pro-Russia slant and low levels of the pro-Ukraine 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 pro-Russia slant, and vary in the amount of articles with the pro-Ukraine slant. Influenced news outlets have few articles that contain pro-Ukraine slant, and vary in the amount of pro-Russia slant.

Results in Subfigure (a) of Figure 10 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 report, 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 pro-Ukraine and pro-Russia slant and the volume of slant as the sum pro-Russia and pro-Ukraine slant.[37] Subfigure (b) presented the resulting

---

[37]We normalize the mean and standard deviations of the measures of the pro-Russia and pro-Ukraine slant

Figure 10: Ideological positions of the news outlets in the Ukraine-crisis news coverage.
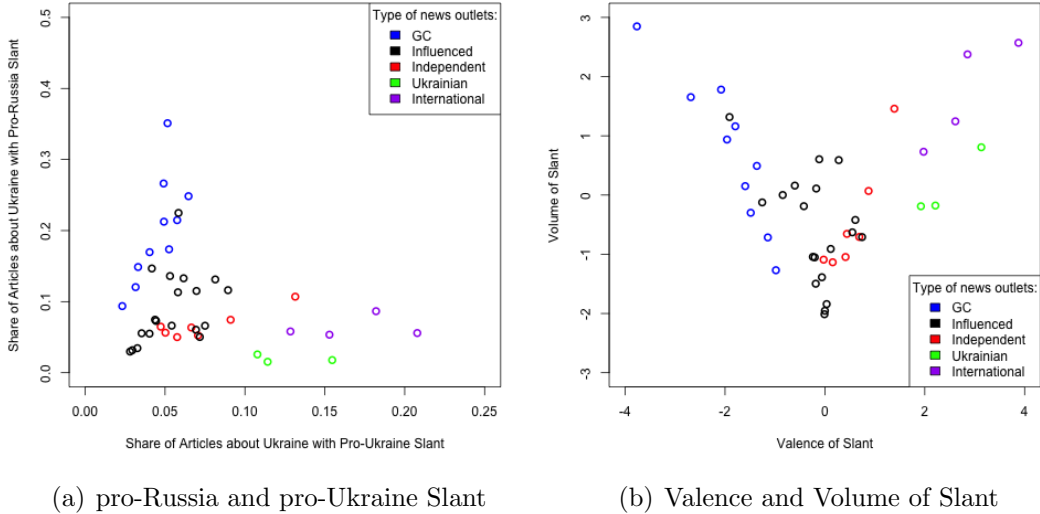
(a) pro-Russia and pro-Ukraine Slant   (b) Valence and Volume of Slant

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.

volume (V_j^+) and valence (V_j^-) of the slant of the news outlets. The GC and some influenced news outlets have negative valence of slant (corresponding to more pro-Russia slant in the reporting), while international, Ukrainian and some independent news outlet have position valence (more pro-Ukraine slant). The majority of independent and influenced news outlet are neutral in terms of the valence of slant and differ in its volume.

4   Descriptive Evidence

In the pervious 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 estimation a demand model for news. However, before we build a demand model, we present some model-free 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 important on the days with more sensitive news. If consumers prefer the pro-government bias, on the days with more sensitive news

_____

to make them comparable.

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 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 internal-sensitive news events and Ukraine-crisis news events on day t:

$$\log(\text{share})_{jt} = b_{0j} + b_j^{IS} log(F_t^{IS}) + b_j^{Ukr} log(F_t^{Ukr}) + X_{jt}'d + \xi_{jt} \qquad (1)$$

where F_t^{IS} and F_t^{Ukr} correspond to the share of articles about internal-sensitive news and Ukraine-crisis news, and X_{jt} correspond to the controls, such as indicator variables for week-days and time trends.[38] The slope coefficients b_j^{IS} and b_j^{Ukr} correspond to the change in the market shares due to the change in the amount of sensitive news in the market.

We estimate b_j^{IS} and b_j^{Ukr} for 42 news outlets including weekday and week indicator variations as controls.[39] Figure 11 summarizes and visualizes the estimation results. Each point on the subfigures (a)-(c) represents an estimate of b_j^{IS} or b_j^{Ukr} for the news outlet j. Points of larger size represent larger absolute value of the estimates, with blue and red colors corresponding to positive and negative estimates of b_j^{IS}, respectively. Points with bold borders represent outlets with statistically significant estimates of b_j^{IS}.[40]

Subfigure (a) of Figure 11 visualizes the estimates of b_j^{IS}. Results suggest that news outlets with higher reporting about the internal-sensitive news are more likely to get an increase in the market shares on the days with more news about the internal-sensitive events. We test this more formally by regressing the b_j^{IS} estimates on $\bar{F}_j^{IS}$, the average share of reporting about the internal-sensitive events by the news outlets j. Table 7 presents the results of this regression based on b_j^{IS} with the different level of controls in regression (1). In the specification with weekday and week fixed effect (column 4) that we've used above,

---

[38]In case of the observations with zero market share, we assign the lowest observed non-zero share of this outlet to this observation.

[39]We exclude 5 news outlets for which we do not have information about the texts of the articles, and 1 news outlets (`dw.de/ru`) for which we have few (10) news consumption occasions.

[40]Significance is tested at 5% level, standard errors are heteroskedasticity and autocorrelation consistent.

Figure 11: Predicted changes in the news outlets' market shares with the change in the amount of sensitive news, by news outlet
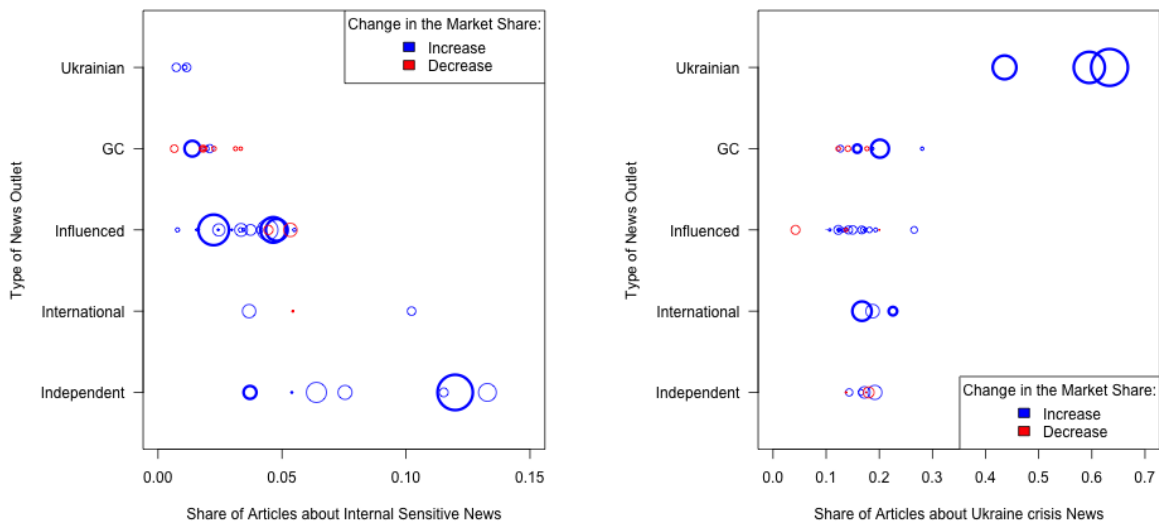
(a) Volume of Internal-sensitive news reporting

(b) Volume of Ukraine-crisis news reporting
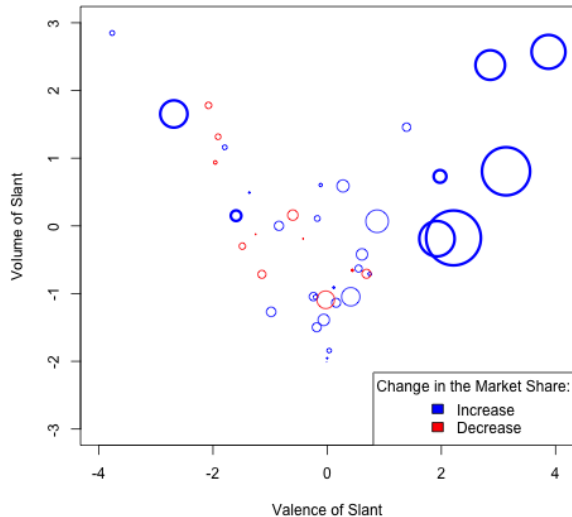
(c) Slant in Ukraine-crisis news reporting

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. Blue color corresponds to the increase in the market shares, and red color corresponds to the decrease in the market share. Bold borders of the points correspond to significance of the change in the market share.

26

the relationship between $\bar{F}_j^{IS}$ and $\hat{b}_j^{IS}$ 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 $\bar{F}_j^{IS}$ and $\hat{b}_j^{IS}$ is significant either on 5% or is on the margin of significance. We interpret this as evidence that news outlets with higher reporting about the internal 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_j^{IS}, and their ideological positions on internal-sensitive news $\bar{F}_j^{IS}$.

	(1) $\hat{b}_j^{IS}$	(2) $\hat{b}_j^{IS}$	(3) $\hat{b}_j^{IS}$	(4) $\hat{b}_j^{IS}$
$\bar{F}_j^{IS}$	0.107^*	0.124^+	0.252^*	0.301^+
	(0.051)	(0.067)	(0.1)	(0.149)
Controls:				
Weekday FE	N	Y	Y	Y
Time trend polynomial (4-order)	N	N	Y	N
Week FE	N	N	N	Y

$^{**} - 1\%$ significance, $^* - 5\%$ significance, $^+ - 10\%$ significance. Controls are included in the regression (1) estimating b_j^{IS}. Standard errors are heteroskedasticity consistent.

Subfigures (b) and (c) of Figure 11 visualize the estimates of b_j^{Ukr}. Results suggest that news outlets with higher reporting about the Ukraine crisis news (subfigure b), less 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_j^{Ukr} estimates on the average share of reporting about the Ukraine-crisis news, $\bar{F}_j^{Ukr}$, 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_j^{Ukr} with the different level of controls in regression (1). Based on the specification with weekday and week fixed effect (column 4) that we've used above, there is a statistical significant positive relationship between $\hat{b}_j^{Ukr}$ and $\bar{F}_j^{Ukr}$, V_j^- and V_j^+, supporting the claim that news outlets that report more about Ukraine crisis, have 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 $\hat{b}_j^{Ukr}$ 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 casual, under the conditional independence assumption (CIA) of the proxy for the level of sensitive news on day t, $\log(\text{share})_{jt} \perp log(F_t^x)|X_{jt} \; \forall j, x = \{IS, Ukr\}$.

Table 8: Relationship between the estimates of the market share changes of news outlets, b_j^{Ukr}, and their ideological positions on Ukraine crisis news, $\bar{F}_j^{Ukr}$, V_j^- and V_j^+.

	(1) $\hat{b}_j^{Ukr}$	(2) $\hat{b}_j^{Ukr}$	(3) $\hat{b}_j^{Ukr}$	(4) $\hat{b}_j^{Ukr}$
$\bar{F}_j^{Ukr}$	1.134*	1.124*	0.699**	0.736**
	(0.480)	(0.481)	(0.213)	(0.119)
V_j^-	−0.010	−0.010	0.055$^+$	0.035**
	(0.056)	(0.057)	(0.030)	(0.011)
V_j^+	0.022	0.021	−0.021	0.024*
	(0.062)	(0.062)	(0.033)	(0.012)
Controls:				
Weekday FE	N	Y	Y	Y
Time trend polynomial (4-order)	N	N	Y	N
Week FE	N	N	N	Y

** − 1% significance, * − 5% significance, $^+$ − 10% significance. Controls are included in the regression (1) estimating b_j^{Ukr}. Standard errors are heteroskedasticity consistent.

This is a plausible assumption given that $log(F_t^x)$ is determined by sensitive news events happening on day t, which is out of control of the market participants. Under the CIA, result 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 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 higher volume of slanted news. Another example is sorting of consumers who prefer the independent news outlets to reading the news on the days with a lot of internal-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-government biased coverage. Second, 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.

To separate out these alternative explanations, in the next section we formulate a structural demand model and use the 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, internal-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_t^x, where x corresponds to non-sensitive, internal-sensitive or Ukraine-crisis news, $x = \{Non, IS, Ukr\}$. We assume that the news-market participants take the underlying production process as given.

As researchers, we do not observe N_t^x. 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), $F_t^x = \frac{\sum_j N_{jt}^x}{\sum_x \sum_j N_{jt}^x}$, where N_{jt}^x 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 $type_j$: independent, influenced, GC, Ukrainian, or international. News outlet j chooses its quality α_j, its level of reporting about the internal-sensitive and Ukraine-crisis news, $\bar{F}_j^{IS}$ and $\bar{F}_j^{Ukr}$, and its valence and volume of slant in the Ukraine-crisis news coverage, V_j^- and V_j^+. 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 τ on day t, consumer i can choose one news outlet, or choose an outside option of not consuming any news.[41] 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 τ is determined by the earliest news article visited by the consumer i on an outlet j on day t.

We assume that consumer have information about the relative importance of news topic 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_j^- and V_j^+ $\forall$ 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, β_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, ρ_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. We define the utility of consumer i from an outlet j on day t and on consumption occasions τ as

$$u_{ijt\tau} = \alpha_{ij} + F_t^{Ukr}\left(\eta_i^{Ukr} + \bar{F}_j^{Ukr}\beta_i^{Ukr} + V_j^-\gamma_i^- + V_j^+\gamma_i^+ + |V_j^- - V_{y_{i\tau-1}}^-|(\tau > 1)\rho_i\right) + \quad (2)$$
$$+ F_t^{IS}\left(\eta_i^{IS} + \bar{F}_j^{IS}\beta_i^{IS}\right) + |V_j^- - V_{y_{i\tau-1}}^-|(\tau > 1)\eta_i^- + s_{it\tau}\eta_i^s + \epsilon_{ijt\tau},$$

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 η_i^{Ukr} and η_i^{IS} explain the changes in the consumer utility due to the changes in the amount of news topics happening on day t.[42] The measure

---

[41]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 consumer to navigate to multiple news outlets on the same day t.

[42]Given that we observe only relative importance of news topics over time, F_t^{Ukr} and F_t^{IS}, we normalize consumer preference for non-sensitive news to zero, $\eta_i^{Non} = \beta_i^{Non} = 0$.

$|V_j^- - V_{y_{i\tau-1}}^-|$ captures the ideological distance in the Ukraine crisis slant between consumer i's current and previous consumption choice (denoted as $y_{i\tau-1}$) on day t, and it does not affect consumer utility on the first consumption occasion on day t. Thus, coefficient η_i^- captures the baseline variety-seeking behavior of consumers and allows to interpret ρ_i as the preference for ideological diversity in the Ukraine-crisis coverage. Finally, state variable $s_{it\tau}$ 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, $s_{it\tau}$ allows us to capture the fact that consumers do not return to an outlet j after the consumption while still keeping it in the potential choice set. The choice of an outside option is defined as being online but not navigating to the news articles, and it is normalized to $u_{i0t\tau} = \epsilon_{i0t\tau}$.

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 few sensitive news ($F_t^{Ukr} = F_t^{IS} = 0$), the news consumption utility for consumer i comes from the fixed preferences of this consumer from the new outlets, α_i. Thus, consumption choices on these days identify α_i. Reduced-form parameters η_i^- and η_i^s are identified from the occasions with multiple news outlets consumed within a day. On the days with more sensitive news ($F_t^{Ukr} > 0; F_t^{IS} > 0$), consumer i derives utility both from the fixed effects α_i, but also from her preferences for the sensitive news topics. Parameters η_i^{Ukr} and η_i^{IS} are identified from the likelihood to consumer any news outlet on the days with more sensitive news. Ideological preferences β_i^{IS}, β_i^{Ukr}, γ_i^-, and γ_i^+ are identified from consumer i's switching on the days with more sensitive news. Finally, preference for the ideological diversity, ρ_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.[43] 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.

Due to the computational limitations, we use a random sample of 10,000 news consumers out of the specified sample of 52,568 news consumers in the estimation procedure.

6.2   Estimation

We estimate the distribution of $\theta_i = \{\alpha_{ij}, \eta_i^{Ukr}, \eta_i^{IS}, \eta_i^-, \eta_i^s, \beta_i^{Ukr}, \beta_i^{IS}, \gamma_i^-, \gamma_i^+, \rho_i\}$ using a Bayesian hierarchical model. We assume that $\epsilon_{ijt\tau}$ follows a type-1 extreme value distribution. At each choice occasion τ 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 at on day t on the consumption occasion τ is

$$\pi(y_{it\tau} = j|\theta_i) = \frac{exp(u_{ijt\tau}(\theta_i))}{1 + \sum_{j'} exp(u_{ij't\tau}(\theta_i))}.$$

The likelihood of θ_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 θ_i is a normal distribution, with the normal prior over its mean and the inverse Wishart prior over the covariance matrix:

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

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

---

[43]Out of 214,375 news consumer who visit a news article page at least once over the sample period.

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

We estimate the distribution of the parameters θ by simulating from the posterior distribution using an MCMC hybrid sampler. We pick standard tuning parameters following Rossi, Allenby, and McCulloch (2005).

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 Ukraine-crisis news to not sensitive news ($E(\hat{\eta}^{Ukr}) = 0.223$) and not sensitive news to the internal-sensitive news ($E(\hat{\eta}^{IS}) = -0.032$), as reflected by both consumer preference for news topics under the average coverage of news outlets ($E(\hat{\eta}^{Ukr}) = 0.223$ and $E(\hat{\eta}^{IS}) = -0.032$) and by preference for higher coverage of sensitive news ($E(\hat{\beta}^{Ukr}) = 0.062$ and $E(\hat{\beta}^{IS}) = -0.006$). However, there is substantial heterogeneity in consumer preferences, with 49.7% of consumers prefer more than average coverage of internal-sensitive news, and 40.99% prefer less then average coverage of the Ukraine-crisis news. Second, an average consumer prefers the Ukraine-crisis news coverage with less pro-government slant ($E(\hat{\gamma}^-) = 0.071$) and less slant in general ($E(\hat{\gamma}^+) = -0.009$), but once again there is significant heterogeneity in consumer preferences, with 39.58% of consumers having a preference for more pro-government slant. Finally, the vast majority of consumers in the sample, 72.24%, prefer to read more ideological-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, α_j, representing the stable characteristics of the news outlets, such as quality. To compare these preferences across the news outlet types, we aggregate α_j by outlet types, $\hat{\alpha}_{\text{type}}$, and demean it by average α_j across all the news outlets, $\hat{\bar{\alpha}}$. An average consumer prefers the GC news outlets the most ($\hat{\alpha}_{GC} - \hat{\bar{\alpha}} = 0.582$), followed by the influenced ($\hat{\alpha}_{Inf} - \hat{\bar{\alpha}} = 0.322$) and independent ($\hat{\alpha}_{Inf} - \hat{\bar{\alpha}} = -0.061$) 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 (88.04% and 83.34%, 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 (77.6%).[44] Second, the

---

[44]0.6% posterior standard deviation.

Table 9: Posterior point estimates of consumer preferences for news coverage.

	Mean	S.D.	% of users > 0
$\hat{\eta}^{IS}$	-0.032	0.182	43.71
	(0.002)	(0.002)	(0.58)
$\hat{\beta}^{IS}$	-0.006	0.155	49.70
	(0.002)	(0.002)	(0.62)
$\hat{\eta}^{Ukr}$	0.223	0.581	65.72
	(0.005)	(0.005)	(0.48)
$\hat{\beta}^{Ukr}$	0.062	0.279	59.01
	(0.013)	(0.007)	(1.93)
$\hat{\gamma}^{-}$	0.071	0.271	60.42
	(0.006)	(0.004)	(0.80)
$\hat{\gamma}^{+}$	-0.009	0.207	48.22
	(0.004)	(0.003)	(0.85)
$\hat{\eta}^{-}$	0.141	0.357	64.21
	(0.004)	(0.004)	(0.44)
$\hat{\rho}$	-0.121	0.202	27.76
	(0.003)	(0.003)	(0.51)

Posterior standard deviation estimate is in the brackets.

majority of consumers get disutility from the ideological positions of the GC news outlets on the Ukraine-crisis news, given that they exhibit more pro-government slant and more slant in general than the independent news outlets. These consumers get disutility from the government control in this market. 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. Moreover, for the internal-sensitive news, where government control affects the ideological positions through censorship, preferences of consumers are roughly split, indicating that a lot of them prefer the ideological position of the GC news outlets. Thus, some consumers might navigate to the GC news outlets because of their ideological positions.

To reconcile this conflicting evidence, we compare the magnitudes of 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_t^{IS} and F_t^{Ukr}, to have a unit mean, and the reporting and slant decisions of the news outlets, $\bar{F}_j^{IS}$, $\bar{F}_j^{Ukr}$, 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

Table 10: Posterior point estimates of persistent preferences for news outlets.

	Mean	S.D.	% of users > 0
$\hat{\bar{\alpha}}$	-6.627	1.253	0.00
	(0.009)	(0.007)	–
$\hat{\alpha}_{GC} - \hat{\bar{\alpha}}$	0.582	0.500	88.04
	(0.007)	(0.006)	(0.42)
$\hat{\alpha}_{Ind} - \hat{\bar{\alpha}}$	-0.061	0.615	46.13
	(0.009)	(0.011)	(0.65)
$\hat{\alpha}_{Inf} - \hat{\bar{\alpha}}$	0.322	0.341	83.34
	(0.008)	(0.006)	(0.69)
$\hat{\alpha}_{Int} - \hat{\bar{\alpha}}$	-1.132	1.241	17.32
	(0.023)	(0.026)	(0.59)
$\hat{\alpha}_{Ukr} - \hat{\bar{\alpha}}$	-2.673	2.318	10.92
	(0.058)	(0.036)	(0.26)

Posterior standard deviation estimate is in the brackets.

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 0.223 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 0.071 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 internal-sensitive news, censorship is the mechanism of government control, so we compare the utilities consumers get from the amount of coverage about internal-sensitive news on day t, $F_t^{IS} \hat{\beta}_i^{IS} (\bar{F}_{GC}^{IS} - \bar{F}_{Ind}^{IS})$. The difference in coverage between the GC and independent news outlets is 2.17 standard deviations, so an average consumers gets $0.006 * 2.17 = 0.013$ more utility from an average GC 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 0.643 between the GC and independent news outlets. Subfigure 12 (a) plots changes in the difference of consumer utilities from an average GC and independent news outlet as the volume of internal-sensitive news increase in the market. Fraction of consumers who prefer an average GC news outlet to the independent news outlet stays almost the same, reducing from 77.6% on days with no sensitive news to 74.2% on days with twice the average of sensitive news, showing that the quality difference is more important for consumers than the ideological difference.[45]

For the Ukraine-crisis news, slant is the mechanism of government control, so we compare

---

[45]Posterior standard deviation estimate of the difference is 0.48%.

Figure 12: Distribution in the expected utility difference between an average GC and independent news outlet.

(a) Internal-sensitive news

(b) Ukraine-crisis news
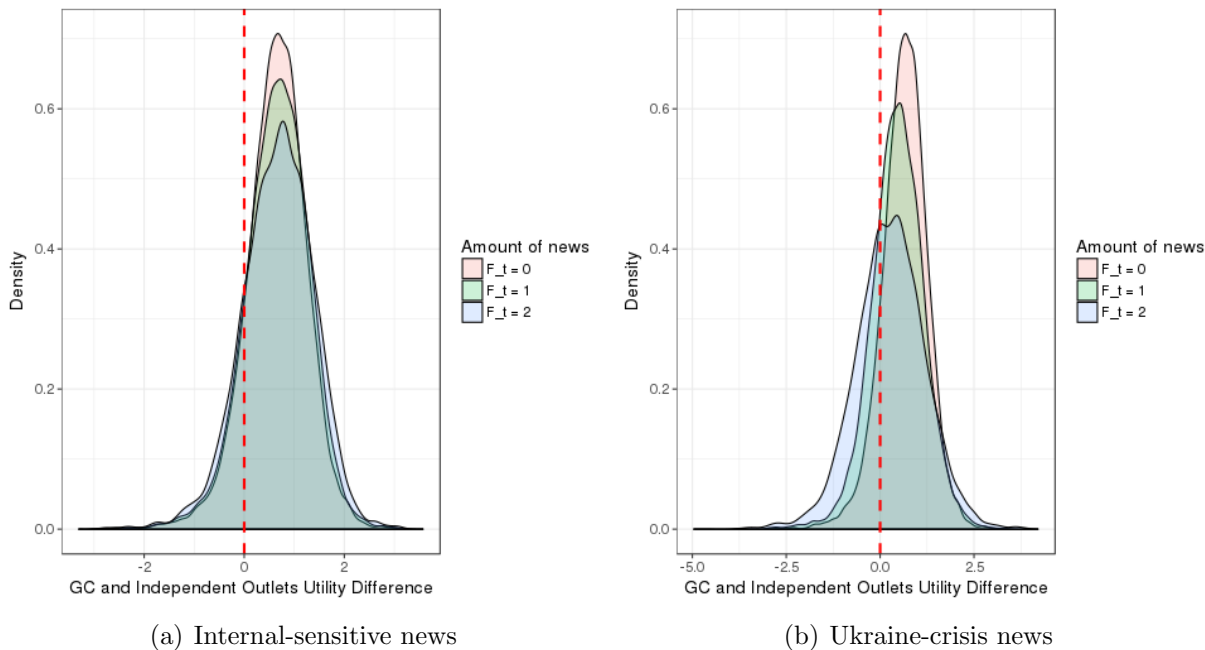
the utilities consumers get from slant in the Ukraine-crisis coverage on day t, $F_t^{Ukr}(\hat{\gamma}^-(V_{GC}^- - V_{Ind}^-) + \hat{\gamma}^-(V_{GC}^+ - V_{Ind}^+)$. An average GC news outlet has 2.45 standard deviations more pro-government slant $(V_{GC}^- - V_{Ind}^- = -2.45)$ and 1.12 standard deviation more slant in general $(V_{GC}^+ - V_{Ind}^+ = 1.12)$ compared to an average independent news outlet, so an average consumer gets $0.071 * 2.45 + (-0.009) * -1.12 = 0.184$ 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 internal-sensitive news, it is still lower than the role of quality. Subfigure 12 (b) plots changes in the difference of consumer utilities from an average GC and independent news outlet as the volume of the Ukraine-crisis news increase in the market. With an increase in the volume of the Ukraine-crisis news, a share of consumer who prefer the GC news outlets fall, with 58.5% (0.56%) 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 consumer 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.

Results above show that the quality difference between the GC and independent news outlets play 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 outlet under different levels of quality and government control. Before we move to the counterfactuals, we discuss three additional characteristics of the consumer preference estimates.

First, the average estimates of consumer preferences do not match the implied consumer preferences from our descriptive analysis. In particular, using only information on the market share levels over time and assuming homogeneity of consumer preferences, we would expect an average consumer to prefer internal-sensitive news to not sensitive news. As is evident from the demand estimates, consumer preferences exhibit a lot of heterogeneity, which makes the market share analysis lead to potentially wrong conclusions. In the Appendix 9.8, we reconcile the descriptive evidence results with the demand estimates, examining changes in the market shares under the estimated preferences. We further show that the marginal distribution of β_i^{IS} drives the difference between the structural estimates and the descriptive evidence, as opposed to other explanations such as the correlation structure between consumer preferences for coverage and news outlets.

Second, given that we find that persistent preferences of consumers play a crucial role in the demand for news, we approach to describe the nature of the persistent preferences of consumers, α_{ij}. While we refer to the persistent preferences of consumers as quality, it potentially includes any characteristics of the website, such as the breadth of news coverage and brand capital. In particular, ideological position of the news outlet might affect persistent preferences of the consumers in the long term 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 9.9 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 β and γ as a short-term effect of the ideological positions on the market shares.

Finally, structural demand estimates allow us to separate out the alternative explanations for the nature of the demand for bias: 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: only 27.76% of consumers start navigating to the ideologically-diverse news outlets on days with more sensitive news. We also conclude that the preferences of a substantial share (48.22%) of the news readers can be explained by

the demand for slant as pure entertainment as they tend to prefer news outlets with more ideological slant in general. However, the data shows that confirmation bias is a 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 play 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, β and γ, and not through the potential changes in the persistent preferences.[46] Throughout the counterfactuals, we treat the average ideological position of the independent news outlets as "unbiased", and deviations from this average position as a results of government control.[47]

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, market share of the GC news outlets decrease by 42.8%, from a 7.25% share to a 4.15% share. The influenced news outlets benefit the most from this reduction in quality as their market share increases by 5.7%. However, most of the switching consumers, 72.2%, choose not to read online news outlet after the quality decrease.

Columns (3)-(6) present the predicted market shares under the counterfactual levels of 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

---

[46]Thus, we do not consider long-term factors like brand capital formation.

[47]Naturally, such 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, estimate the costs parameters and 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.

Table 11: Simulated market shares for different levels of government control and quality of the GC news outlets.

Market shares (%)	(1) Actual	(2) Low α_{GC}	(3) Direct	(4) No control Indirect	(5) Both	(6) More control
$\bar{sh}_{Gov}$	7.25	4.15	8.48	7.20	8.31	7.27
	(0.02)	(0.01)	(0.06)	(0.02)	(0.05)	(0.02)
$\bar{sh}_{Inf}$	10.01	10.58	9.79	10.54	10.21	10.05
	(0.02)	(0.02)	(0.02)	(0.05)	(0.04)	(0.02)
$\bar{sh}_{Ind}$	6.26	6.50	6.10	6.16	6.04	5.99
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
$\bar{sh}_{Int}$	0.60	0.63	0.58	0.59	0.57	0.61
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\bar{sh}_{Ukr}$	0.98	1.01	0.97	0.98	0.97	0.98
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\bar{sh}_{\text{Outside}}$	74.90	77.14	74.08	74.54	73.89	75.10
	(0.04)	(0.03)	(0.05)	(0.05)	(0.07)	(0.04)

The market share are in percent of the entire market. Posterior standard deviation estimate is in the brackets.

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, $F_j^{IS,\text{new}} = F_j^{IS} - \frac{\sum_{j' \in GC} F_{j'}^{IS}}{\sum_{j' \in GC} 1} + \frac{\sum_{j' \in Ind} F_{j'}^{IS}}{\sum_{j' \in Ind} 1}$, $V_j^{-,\text{new}} = V_j^{-} - \frac{\sum_{j' \in GC} V_{j'}^{-}}{\sum_{j' \in GC} 1} + \frac{\sum_{j' \in Ind} V_{j'}^{-}}{\sum_{j' \in Ind} 1}$ and $V_j^{+,\text{new}} = V_j^{+} - \frac{\sum_{j' \in GC} V_{j'}^{+}}{\sum_{j' \in GC} 1} + \frac{\sum_{j' \in Ind} V_{j'}^{+}}{\sum_{j' \in Ind} 1}$ $\forall \ j \in GC$. Without the direct control, market shares of the GC news outlets increase by 17%, with most of the traffic coming from the extrinsic 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.[48] The total advertising expenditure in Russian internet on display advertising in 2014 was 19.1 billion rubles[49], which is around $318 million using the 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 1.23 percentage points reduction

---

[48]Only one of the news outlets in the sample, slon.ru, used paid subscription starting in 2015.
[49]http://www.akarussia.ru/knowledge/market_size

in the market share of the GC news outlets due to direct control is small, corresponding to $\frac{1.23}{25.1} * \$318 = \$15.7$ million dollars. For comparison, government subsidies to mass media in Russia in 2015 were 72.6 billion rubles (\$1.21 billion), which is around 77 times higher than the advertising loss.[50]

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 5.02% of their market share. Results in column (5) show that under no control (direct or indirect) GC and influenced news outlets both gain higher market shares, although an increase in their market shares is smaller than if only direct or indirect control is removed.[51]

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.[52] Independent news outlets lose 4.3% of their market share if their average ideological position is matched to 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 \$3.4 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 outlets. 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

---

[50]Source: http://www.rbc.ru/politics/29/06/2015/55912ffa9a7947453982cda9. Same exchange rate used. The total of 72.6 billions rubles includes subsidies to the television and print.

[51]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 ($7.25 + 10.01 = 17.26\%$) compared to the regimes with only direct and only indirect control ($7.2 + 9.79 = 16.99\%$).

[52]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) and change it's ownership later in 2017 (http://www.forbes.ru/milliardery/346333-berezkin-kupil-u-prohorova-rbk).

day t to an outlet j as

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

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

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

Using this definition, attention share of the GC news outlets is 33.1% (0.1%), corresponding to the media power of 0.5 under the worst-case scenario assumptions.[53] Media power of 0.5 allows the government to swing 75-25% elections into a draw.

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 21.98% (0.09%), corresponding to 0.28 media power. Such media power allows the government to swing 64-36% 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:

$$\Pr(y_{it} = j|\Delta U_i^x < 0)/\left(1 - \Pr(y_{it} = 0|\Delta U_i^x < 0)\right),$$

where ΔU_i^x is the utility consumer i gets from the pro-government bias in sensitive news x topic, $\Delta U_i^{IS} = \hat{\beta}_i^{IS}(\bar{F}_{GC}^{IS} - \bar{F}_{Ind}^{IS})$ and $\Delta U_i^{Ukr} = \hat{\gamma}^-(V_{GC}^- - V_{Ind}^-) + \hat{\gamma}^-(V_{GC}^+ - V_{Ind}^+)$. We use this measure to compute the attention share of the GC news outlets over consumers with $\Delta U_i^x < 0$ 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 shares are 29.7% for a big internal-sensitive news day and 19.7% for a big Ukraine-crisis news day.[54] Under the lower quality of the GC news outlets, the attention shares on such days change to 19.3% and 12.2%, respectively. Thus, high quality of the GC news outlets allow them to capture an additional 7.5%-10.5% of consumers who prefer the ideological coverage of the independent news outlets.

---

[53]Including that the readers are naive and do not understand that the GC news outlets are trying to persuade them. For more details, please see Prat (2017) and Kennedy and Prat (2017).

[54]Big sensitive news day is a day with three times the average amount of sensitive news.

8   Conclusion

In the new era of broad access to information, it is critical to understand whether and how governments can control the 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 substantial market share even though the majority of the population dislike their ideological bias. The main driver for this 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 43% of the controlled news outlets' market shares and 33% of their attention share.

To address this question, we first characterize the reporting of the news outlets using their publications records. We show that there are multiple ways in which the coverage of the GC news outlets is different from the independent outlets, including omission of certain topics (censorship) and distortion of sensitive news topics (propaganda). We use these differences to describe the ideological positions of the news outlets and measure the amount of sensitive news that happened over time.

We then build and estimates a demand model for news, in which we disentangle persistent preferences of consumers for news outlets and their preferences for the ideological position in the sensitive news coverage. We separately identify these preferences using changes in the amount of sensitive news that happen over time, with news consumption information coming from a detailed browsing panel, Internet Explorer Toolbar data.

Estimation results reveal that, while there is heterogeneity in consumer preferences, the majority of consumers have a distaste for the ideological bias of the GC news outlets and a preferences for their quality and brand. If the GC news outlets match the ideological position of the independent outlets, they would get 17% higher market shares, corresponding to a rough back-of-the-envelope estimate of $15.7 million in the advertising revenues. In contrast, if the average quality of the GC news outlets was similar to the independent news outlets, they would get 42.8% lower market share ($39.5 million). In addition to this, high quality of the GC news outlets increase the their share of online attention from 21.98% to 33.1%, substantially increasing the media power. On the days with a lot of sensitive news, such "brand media power" allows the GC news outlets to capture 19.7%-29.7% 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 the alternative explanations for the nature of the demand for bias. We find that the majority of consumers prefer news

about the Ukraine crisis with lower pro-government valence (60.42%) and lower volume of slant (51.78%). The vast majority of consumers, 72.24%, prefer more ideologically-similar news sources on days with more Ukraine-crisis news. Thus, only a minority (27.76%) 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.

We notice, 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 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

Athey, Susan, Markus Mobius and Jeno Pal (2017): "The Impact of Aggregators on Internet News Consumption," Working paper.

Anderson, Simon, David Strömberg and Joel Waldfogel (2016): Handbook of Media Economics 1st Edition, North Holland.

Bai, Jie, Mikhail Golosov, Nancy Qian and Yan Kai (2014): "Understanding the Influence of Government Controlled Media: Evidence from Air Pollution in China," Working Paper.

Baron, David (2006): "Persistent Media Bias," *Journal of Public Economics*, 90: 1-36.

Besley, Timothy and Robin Burgess (2002): "The Political Economy of Government Responsiveness: Theory and Evidence from India," *Quarterly Journal of Economics*, 117: 1415-51.

Besley, Timothy and Andrea Prat (2006): "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability," *Americal Economic Review*, 96: 720-736.

Cagé, Julia, Nicolas Hervé and Marie-Luce Viaud (2017): "The Production of Information in an Online World: Is Copy Right?," Working paper.

Della Vigna, Stefano and Ethan Kaplan (2007): "The Fox News Effect: Media Bias and Voting," *Quarterly Journal of Economics* 122: 1187-1234.

Dubé, Jean-Pierre, Günter Hitsch and Peter Rossi (2010): "State Dependence and Alternative Explanations for Consumer Inertia," *RAND Journal of Economics*, 41: 417-445.

Durante, Ruben and Brian Knight (2012): "Partisan control, media bias, and viewer responses: Evidence from Berlusconi's Italy," *Journal of the European Economic Association*, 10: 451-481.

Enikolopov, Ruben, Maria Petrova and Ekaterina Zhuravskaya (2011): "Media and Political Persuasion: Evidence from Russia," *American Economic Review*, 101: 3253-85.

Garcia-Arenas, Javier (2016): "The impact of free media on regime change: Evidence from Russia," Working Paper.

Franceschelli, Ignacio (2011): "When the Ink is Gone: The Transition from Print to Online Editions," Working Paper.

Gehlbach, Scott and Konstantin Sonin (2014): "Government Control of the Media," *Journal of Public Economics*, 118: 163-171.

Gentzkow, Matthew, Michael Sinkinson, and Jesse M. Shapiro (2011): "The Effect of Newspaper Entry and Exit on Electoral Politics," *American Economic Review* 101: 2980-3018.

Gentzkow, Matthew and Jesse M. Shapiro (2006): "Media Bias and Reputation," *Journal of Political Economy*, 114: 280-316.

Gentzkow, Matthew and Jesse M. Shapiro (2010): "What Drives Media Slant? Evidence from U.S. Daily Newspapers," *Econometrica* 78: 35-71.

Gentzkow, Matthew and Jesse M. Shapiro (2015): "Ideology and Online News," NBER Working Paper 19675.

Gentzkow, Matthew, Jesse M. Shapiro and Matt Taddy (2016): "Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech," NBER Working Paper 22423.

Kennedy, Patrick and Andrea Prat (2017): "Where do People Get Their News?" Working Paper.

Knight, Brian and Ana Tribin (2016): "The Limits of Propaganda: Evidence from Chavez's Venezuela" NBER Working Paper 22055.

Martin, Gregory J. and Ali Yurukoglu (2015): "Bias in Cable News: Persuasion and Polarization," Working paper.

McCombs Maxwell E. , and Donald L. Shaw (1972): "The Agenda-Setting Function of Mass Media," *Public Opinion Quaterly* 36: 176-187.

Mullainathan, Sendhil and Andrei Shleifer (2002): "Media Bias," NBER Working Paper 9295.

Mullainathan, Sendhil and Andrei Shleifer (2005): "The Market for News," *American Economic Review* 95: 1031-1053.

Pan, Jennifer and Yiqing Xu (2017): "China's Ideological Spectrum," *The Journal of Politics*, Forthcoming.

Perego, Jacopo and Sevgi Yuksel (2016): "Media Competition and the Source of Disagreement," Working paper.

Prat, Andrea and David Strömberg (2013): "The Political Economy of Mass Media," Advances in Economics and Econometrics: Volume 2, Applied Economics.

Prat, Andrea (2017): "Media Power," *Journal of Political Economy*, Forthcoming.

Roberts, Margaret E. (2014): "Fear or Friction? How Censorship Slows the Spread of Information in the Digital Age," Working Paper.

Rossi, Peter (2014): "Bayesian Non- and Semi-parametric Methods and Applications," Princeton University Press.

Rossi, Peter, Greg Allenby and Robert McCulloch (2005): "Bayesian Statistics and Marketing," Wiley Series in Probability and Statistics.

Sen, Ananya and Pinar Yildirim (2016): "Clicks Bias in Editorial Decisions: How does Popularity Shape Online News Coverage?," Working paper.

Xiang, Yi and Miklos Sarvary (2007): "News Consumption and Media Bias," *Marketing Science*, 26(5): 611-628.

Zhu, Yi and Anthony Dukes (2015): "Selective Reporting of Factual Content by Commercial Media," *Journal of Marketing Research*, 52(1): 56-76.

9 Appendices

9.1 News Outlets Classification

News outlets classification was done based on interviews with media professionals and ownership structure of the news outlets. 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 an independent editor-in-chief to a more government-loyal editor-in-chief in September 2013; *lenta* got a similar change in March of 2014.[55]

- *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.[56]

- *kommersant* is owned by Alisher Usmanov, one of the richest Russian oligarchs.[57]

- *kp* is owned by Grigory Berezkin, who is on board of directors of state-owned RZD.[58]

- *fontanka* is owned by "Azur-Media."

---

[55]https://meduza.io/feature/2016/05/17/12-redaktsiy-za-pyat-let
[56]http://www.kommersant.ru/doc/2311510
[57]https://lenta.ru/lib/14164974/
[58]http://www.forbes.ru/profile/grigorii-berezkin

Potentially government-influenced news outlets:

- *bfm* is owned by Rumedia, a company of Russian steel tycoon Vladimir Lisin.[59]

- *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, it is reported to be influenced by the government. Reported to publish paid articles.[60]

- *interfax*'s beneficiary is not disclosed, but there is information that it is owned by the top-management.[61]

- *mk* is owned by Pavel Gusev, a confidant of Vladimir Putin. There are examples of mk removing published articles about government-sensitive topics.[62]

- *znak* was formerly ura.ru, It had to change its name due to government pressure.[63]

- *ng* is owned by Konstantin Remchukov. It is reported to publish articles which are paid for by the government.[64]

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

- *regnum* is reported to have been purchased by Gazprom media.[65] It is reported to publish paid articles.[66]

- *rosbalt, sobesednik* and *trud* are reported to publish paid articles.[67]

Independent news outlets:

- *newsru* is owned by Vladimir Gusinsky, a tycoon opposing 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%).

---

[59]https://en.wikipedia.org/wiki/Vladimir_Lisin
[60]https://tjournal.ru/p/media-denim
[61]https://www.vedomosti.ru/business/articles/2012/01/19/lgota_dlya_smi
[62]http://www.rbc.ru/politics/27/12/2013/897386.shtml
[63]http://www.forbes.ru/news/227994-redaktsiya-uraru-budet-vypuskat-internet-gazetu-znakcom
[64]http://theins.ru/politika/6015
[65]https://lenta.ru/news/2014/06/20/media/
[66]https://tjournal.ru/p/media-denim
[67]https://tjournal.ru/p/media-denim

- *rbc* and *snob* are owned by Mikhail Prokhorov, a Russian billionaire and politician. He participated in the presidential elections of 2012. RBC.ru stayed independent till May 2016, when the top managers were fired due to political pressure.[68] It was later acquired by Grigory Berezkin, 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 air by major TV providers after covering the street protests of 2011. Its website operates based on subscriptions.

- *vedomosti* was jointly owned by Sanoma Independent Media (33%), Financial Times (33%) and The Wall Street Journal (33%) before 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, which is registered in The Netherlands.

International News Outlets:

- *bbc* is a Russian version of BBC.

- *svoboda* is Radio Liberty.

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

- *dw* is a Russian version of Deutsche Welle.

- *reuters* is a Russian version of Reuters.

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

---

[68]http://www.newyorker.com/news/news-desk/the-demise-of-rbc-and-investigative-reporting-in-russia

9.2 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 was scraped directly from the corresponding news websites. On 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 these 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 words that frequently (more than 50% of times used in the corpus) start with a capital letter in the text when they are not in the beginning of the sentence.[69]

---

[69]This way, we include the typical proper nouns but exclude words that are used as proper nouns rarely and only in a certain context.

9.3 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.[70] We classify the URL as a subdirectory if the visited URL matches the subdirectory URL.[71] 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 12 shows statistics of browsing of the 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 12: Summary of browsing behavior

	Page views	Visits (Sessions)	Seconds spent	
			Mean	Median
Main page	5,344,041	1917206	128	42
News articles	1,042,0780	4240831	186	86
News subdirectories	4,225,221	1484410	263	90
Other	6,547,225	2,389,635	145	44
Total	26,537,267	6,630,400	176	64

---

[70]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 ULRs with the structure `http://www.x1.ru/news/topic/year/month/date/some-other-name-of-the-article.html` as news articles.

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

## 9.4 Comparing Weekly Visitors of IE Toolbar and LI

Table 13 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.[72] 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, social network popular with older audience, and less likely to visit vkontakte.ru, a social network popular with younger audience. 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 13: Comparison of other website rankings in IE Toolbar and LI.ru

Website	Description	Visit Share	
		liveinternet.ru	IE Toolbar
auto.ru	Buy/Sell used cars	0.0141	0.0130
avito.ru	Classified posts	0.0701	0.0713
drom.ru	Website about cars	0.0170	0.0215
gismeteo.ru	Weather	0.0347	0.0536
hh.ru	Job postings	0.0138	0.0149
kinopoisk.ru	Movie descriptions	0.0229	0.0109
ngs.ru	Novosibirsk city website	0.0155	0.0067
odnoklassniki.ru	Social Network (older audience)	0.2592	0.4455
pluso.ru	Records clicks to social media	0.1186	0.0000
rutracker.org	Torrent website	0.0182	0.0057
tiu.ru	Online retailer	0.0140	0.0089
vkontakte.ru	Social Network (younger audience)	0.3755	0.3248
wildberries.ru	Online retailer	0.0137	0.0151
woman.ru	Online magazine	0.0129	0.0081

---

[72]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 13.
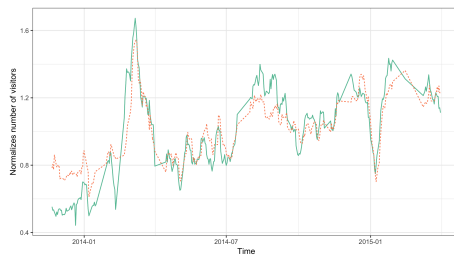
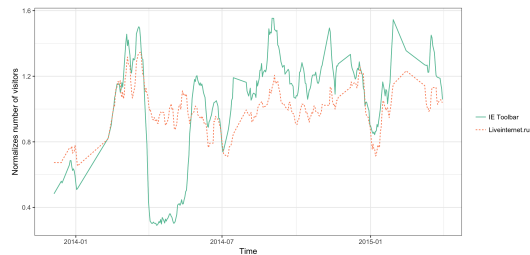Figure 13: Normalized traffic of the top seven news websites, IE Toolbar and Liveinternet.ru
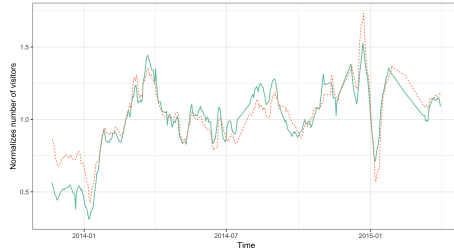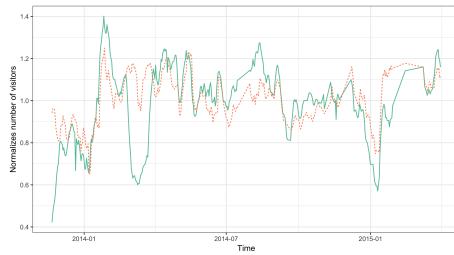

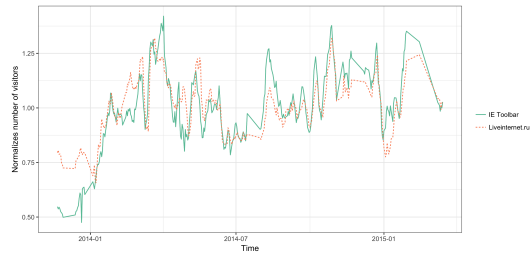(a) ria.ru (cor = 0.914)
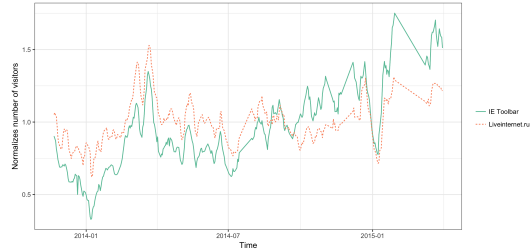
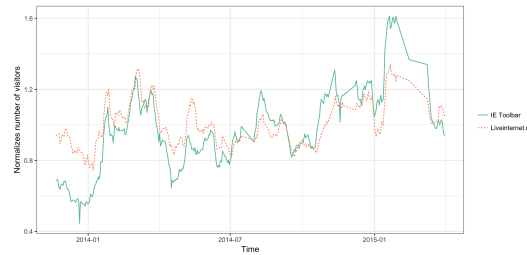(b) ria.ru (cor = 0.702)

(c) lenta.ru (cor = 0.913)

(d) gazeta.ru (cor = 0.520)

(e) vesti.ru (cor = 0.549)

(f) rg.ru (cor = 0.830)

(g) kp.ru (cor = 0.807)

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

9.5 Weekly Users of IE Toolbar

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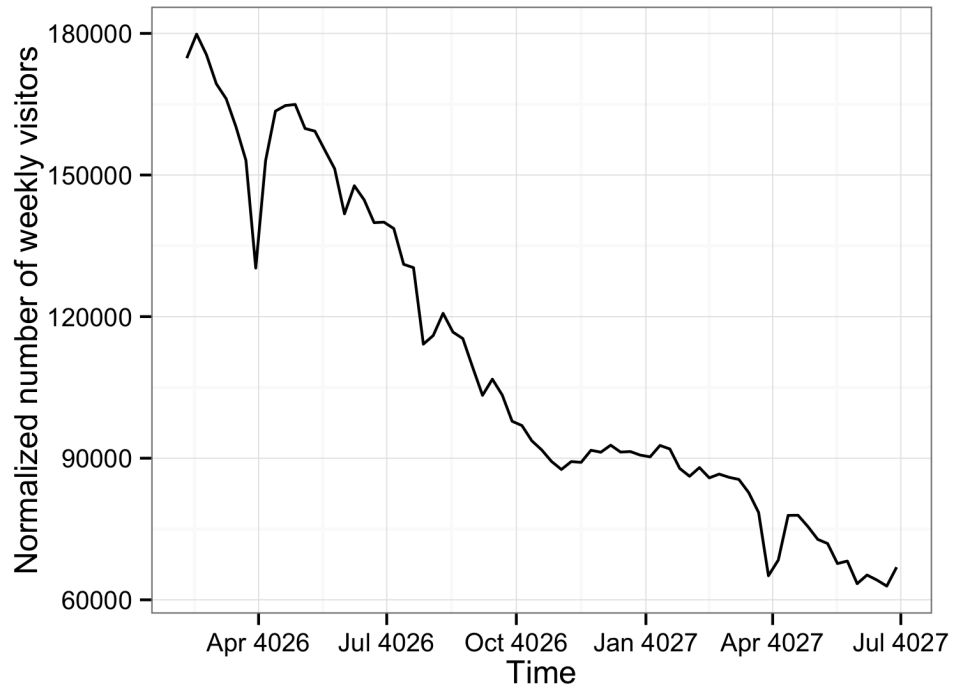
9.6 Sensitive News: Censorship and Slant

9.6.1 Censored unigrams and bigrams

Tables 14 and 15 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 14 and 15).

Table 14: List of the top 54 bigrams of the proper nouns underused by the GC news outlets. Part 1.

Underused proper noun: English translation	Information about the proper nouns	Rank Difference, $\Delta\text{Rank}_v^{Ind-Gov}$
Alexei Navalny	Opposition politician	-28.3
(The) New Times	News outlet	-27.1
Mikhail Khodorkovsky	Opposition politician, political prisoner	-26.7
Echo (of) Moscow	News outlet	-26.6
Dmitry Kiselyov	Journalist	-26.3
Sergei Guriev	Economist, interrogated about "Yukos"	-25.8
Gennady Timchenko	Businessman, friend of Vladimir Putin	-25.7
Galina Timchenko	Journalist	-25.1
Svetlana Davydova	Civilian investigated for treason	-24.6
Alexander Plushev	Journalist	-24.4
Marat Gelman	Gallerist	-24.4
Alexei Navalny (2)	Opposition politician	-24.3
Ilya Yashin	Opposition politician	-24
Pussy Riot	Protest punk rock band	-23.2
Sergey Parkhomenko	Political journalist	-22.9
Alexei Venediktov	Editor-in-Chief of a News Outlet	-22.8
Alexander Vinokurov	Owner of multiple news outlets	-22.3
Arkady Rotenberg	Businessman, friend of Vladimir Putin	-22.3
Andrei Zubov	History professor	-22.2
Mikhail Kosenko	Political prisoner, Bolotnaya protests	-22.1
Alexei Kudrin	Politician, former minister	-21.9
The New (Times)	News outlet	-21.8
Igor Sechin	Chairman of Rosneft, close ally of Putin	-21.8
Ramzan Kadyrov	Head of the Chechen Republic	-21.5
(The) Other Russia	Opposition political party	-21.4

Bigrams marked as bold are selected to define sensitive news.

Table 15: List of the top 54 bigrams of the proper nouns underused by the GC news outlets. Part 2.

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

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 defined 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 the rank difference below the threshold of -21.1. A lot of these unigrams correspond to the last names of the sensitive actors classified based on bigrams, and some other refer to the ambiguous actors.

Figure 15: Histograms of $\Delta\text{Rank}_v^{Ind-Gov}$ across the proper nouns: actual and random corpus.
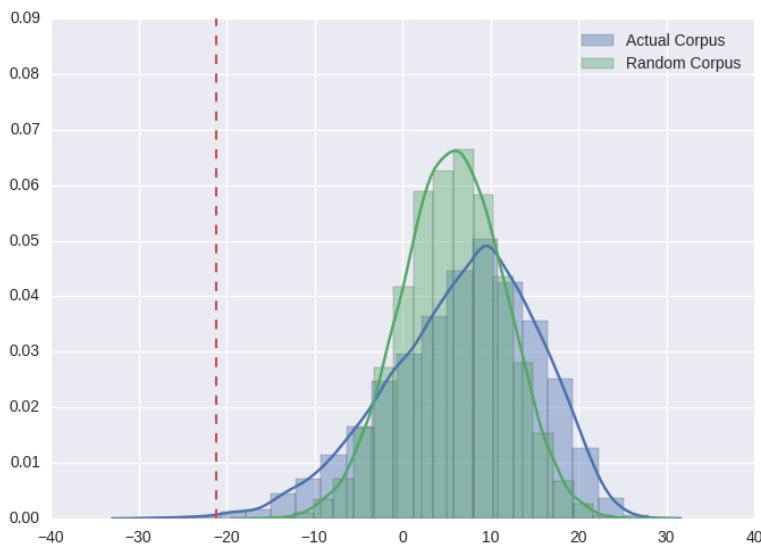
Histogram in blue color corresponds to the actual corpus, histogram in green color – to the random corpus. Red vertical line is a cutoff corresponding to the lowest rank difference in the random sample, -21.1.

Table 16 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 16).

Table 16: List of the top 20 unigrams of the proper nouns underused by the GC news outlets.

Underused proper noun: English translation	Information about the proper nouns	Rank Difference, $\Delta\mathrm{Rank}_v^{Ind-Gov}$
Venediktov		-29.3
Rotenberg		-29
Timchenko		-28.2
Slon	News outlet	-28.1
Revzin	Journalist	-27.9
Roskomnadzor	Federal agency overseeing media	-27.5
Khodorkovsky		-27.4
Venediktov		-27.2
Navalny		-26.4
Plushev		-25.7
Ketchum	PR agency of Russian government	-25.7
Echo		-25.6
Lebedev		-25.5
Kudrin		-25.1
Sechin		-24.9
Kosenko		-24.3
Bolotnaya	Square where protests take place	-24.3
Prohorov		-24.3
Shlosberg	Opposition Politician	-24.2
Sakharov	Ambiguous, might be multiple actors	-24.2
Bukovsky	Ambiguous, might be multiple actors	-23.9
Gelman		-23.8

Unigrams marked as bold are selected to define sensitive news.

9.7 Media Slant in the Ukraine Crisis News

Table 17: List of the words corresponding to the pro-Russia and pro-Ukraine slant.

Overused words by the GC news outlets			Overused words Ukrainian news outlets		
Word	$\Delta\text{Rank}_v^{Ukr-Gov}$	Rank	Word	$\Delta\text{Rank}_v^{Ukr-Gov}$	Rank
reunion	-34.67	1	annexation	-30.8	2
radical	-34.10	2	anti-terrorist	-29.9	4
punitive	-33.47	3	occupied	-29.3	9
overturn	-33.07	4	anti-terrorist (2)	-28.8	18
blockade	-32.60	5	pseudo-referendum	-28.7	20
bombing	-32.20	6	separatist	-28.5	24
coup	-31.73	7	annexed	-28.1	29
anti-Russian	-31.10	8			
bombing (2)	-30.80	12			
russophobe	-30.57	15			
ultra-nationalist	-30.53	16			
neo-nazi	-30.47	18			
intra-Ukrainian	-30.13	20			
nazism	-30.03	23			
russophobe (2)	-28.33	41			
nazi	-27.50	53			
reunion (2)	-27.33	60			
neo-nazi (2)	-27.27	64			

9.8 Preferences Heterogeneity and Predicted Choice Probabilities

9.8.1 Choice Probabilities under Actual Preferences

To get a better understanding of the importance of the amount of sensitive news and news coverage, we simulate the market shares for the news outlets under different levels of sensitive news. Figure 16 presents the predicted news outlet choice probabilities conditional on reading the news, $\Pr(y_i = j|y_i \neq 0)$, on a day with no sensitive news. On average, GC and influenced news outlets have higher choice probabilities compared to independent, international and Ukrainian news outlets. The only notable exception, `rbc.ru`, is an independent news outlet and one of the market leaders.

Figure 16: Posterior estimates of the mean and 95% credibility interval of $\Pr(y_i = j|y_i \neq 0)$ on a day with no sensitive news.
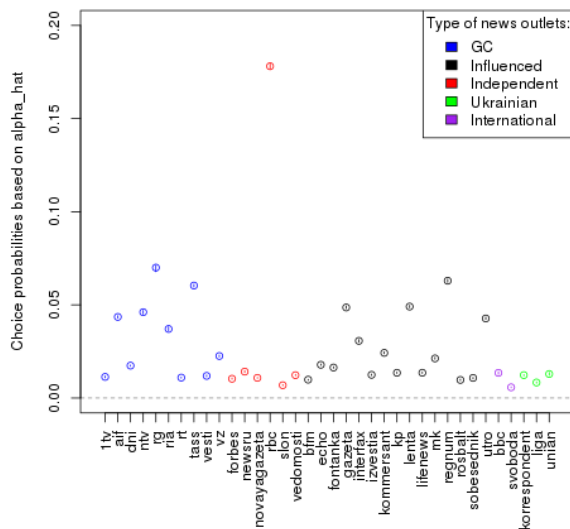

News outlets on the x axis are grouped by their types. Dots represent mean posterior estimate of the choice probabilities. Solid lines represents 95% credibility interval of the estimates.

Figure 17 presents changes in the market shares of the news outlets on days with more internal-sensitive news. Subfigure (a) corresponds to changes in the market shares on a day with an average volume of sensitive news. We conclude that changes in the market shares of the news outlets on these days are relatively small. For example, GC news outlet on average lose 2.7% of the market share, and independent news outlet on average lose around 2% of their market share. Things are different on the days with a lot of sensitive news, presented in Subfigure (b). News outlet that report a lot on internal-sensitive news benefit up to 55%

of their market share. On average, GC news outlet lose around 4.7% of their market share, while independent outlets gain around 16.9%.

Figure 17: Change in predicted choice probabilities, $\Pr(y_i = j)$, on days with (a) an average and (b) three times the average amount of internal-sensitive news observed in the sample.
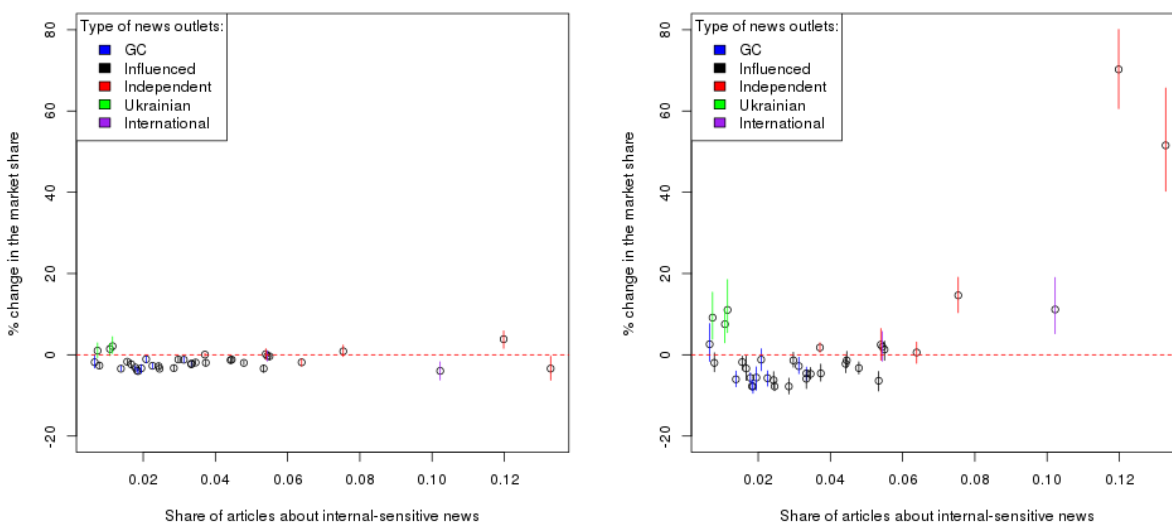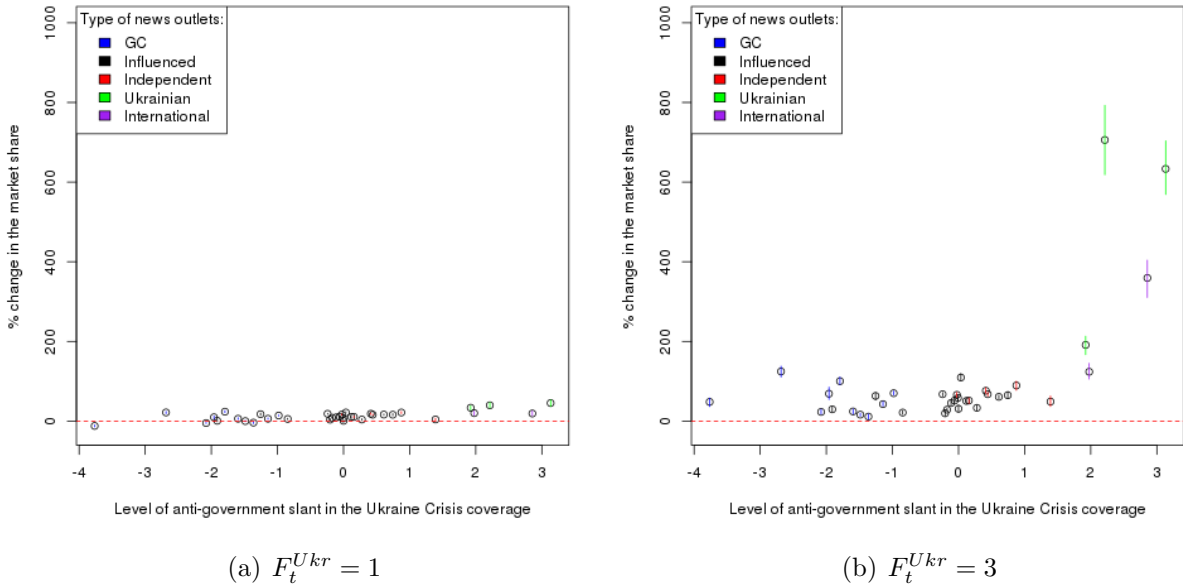

(a) $F_t^{IS} = 1$ (b) $F_t^{IS} = 3$

Dots represent the % change in the market shares of the news outlets on the days with internal-sensitive news as compared to the days with no internal-sensitive news. News outlets are sorted by the share of articles about internal-sensitive news ($\bar{F}_j^{IS}$). Solid lines represents 95% credibility interval of the estimates.

Figure 18 tracks similar information for different amount of Ukraine crisis news. Subfigure (a) shows changes in the market shares on a day with the average amount of news about the Ukraine crisis, compared to a day with no sensitive news. News outlets that have the least pro-government slanted news gain the most, with Ukrainian news outlets gaining around 34% of their market share, followed by the international (19.7%) and independent (16%) news outlets. GC news outlet gain only 6.27% of their market share, with the most pro-government-slanted outlets losing up to 13% of their market share. Once again, results are different on a day with a lot of Ukraine crisis news. Subfigure (b) shows changes in the choice probabilities on a day with 3 times the average amount of sensitive news. News outlets with anti-government slant, such as Ukrainian and international news outlets, gain up to 627% in their choice probabilities. GC news outlets also benefit from higher amount of sensitive news, on average getting 49.5% more visitors.

Figure 18: Change in predicted choice probabilities, $\Pr(y_i = j)$, on days with (a) an average and (b) three times the average amount of internal-sensitive news observed in the sample.

(a) $F_t^{Ukr} = 1$

(b) $F_t^{Ukr} = 3$

Dots represent the % change in the market shares of the news outlets on the days with internal-sensitive news as compared to the days with no internal-sensitive news. News outlets are sorted by the level of anti-government slant in the Ukraine news coverage (V_j^-). Solid lines represents 95% credibility interval of the estimates.

9.8.2 Alternative Heterogeneity Mechanisms

To understand the mechanism behind changes in the choice probabilities we examine consumer preference estimates more closely. First, we examine the role of preference heterogeneity overall. Figure 19 presents changes in the choice probabilities with an increase in the volume of internal-sensitive news under homogenous preferences for internal-sensitive news coverage, $\hat{\beta}_i^{IS} = \hat{E}_i(\beta_i^{IS})$. As expected, with the increase in the amount of internal-sensitive news in the market choice probabilities of the news outlets that report a lot of internal-sensitive news decrease. For example, on a day with 3 times the average amount of internal-sensitive news consumers are 1.43% less likely to read news from independent news outlets, with some news outlets losing up to 13% of their market share. Such decrease is driven by a negative average preferences for internal-sensitive news coverage, $\hat{E}(\beta_i^{IS}) = -0.006$. Such loss in the market share comes in contrast with the results under heterogeneous preferences, where independent news outlets on average gain 16.9% of their market shares on the same day.

Similarly, Figure 20 examines changes in the choice probabilities of consumers under homogenous preferences for slant in the Ukraine-crisis news, $\hat{\gamma}^- = \hat{E}_i(\gamma^-)$ and $\hat{\gamma}^+ = \hat{E}_i(\gamma^+)$. With an increase in the amount of Ukraine-crisis news all news outlets gain in the choice probabilities, and the least pro-government-slanted news outlets gain the most. For example, on a day with 3 times the average amount of Ukraine-crisis news consumers are 27% more likely to read the news from the GC outlets and 540% more likely to read the news from the Ukrainian outlets compared to a day with no Ukraine-crisis news. Thus, qualitatively changes in the choice probabilities are similar to the results under heterogeneous preferences.

Results above show that consumer heterogeneity plays an important role in this market. Independent news outlets gain market share on the days with a lot of internal-sensitive news even though an average consumer dislikes internal-sensitive news topics. One potential mechanism that might affect changes in the market shares is the change in composition of consumers in the market because of sorting. If consumers who prefer independent news outlets also tend to prefer internal-sensitive news, on the days with a lot of internal-sensitive events these consumers will be more likely to read the news, so there will be more readers of independent news outlets in the market. To test for this, we first examine correlation between consumers preferences for independent news outlets, $\alpha_{Ind} - \bar{\alpha}$, and internal-sensitive news, η^{IS}. We find that there is a significant positive correlation of in consumers preferences, $(cor(\alpha_{Ind} - \bar{\alpha}, \eta^{IS}) = 0.071)$.[73] To test whether sorting plays an important in the market, we recompute changes in the choice probabilities under homogenous consumers preferences for

---

[73]Standard deviation across the MCMC draws is 0.0312.

Figure 19: Change in predicted choice probabilities, $\Pr(y_i = j)$, on days with (a) an average and (b) three times the average amount of internal-sensitive news observed in the sample under homogenous preferences for internal-sensitive coverage, β_i^{IS}.
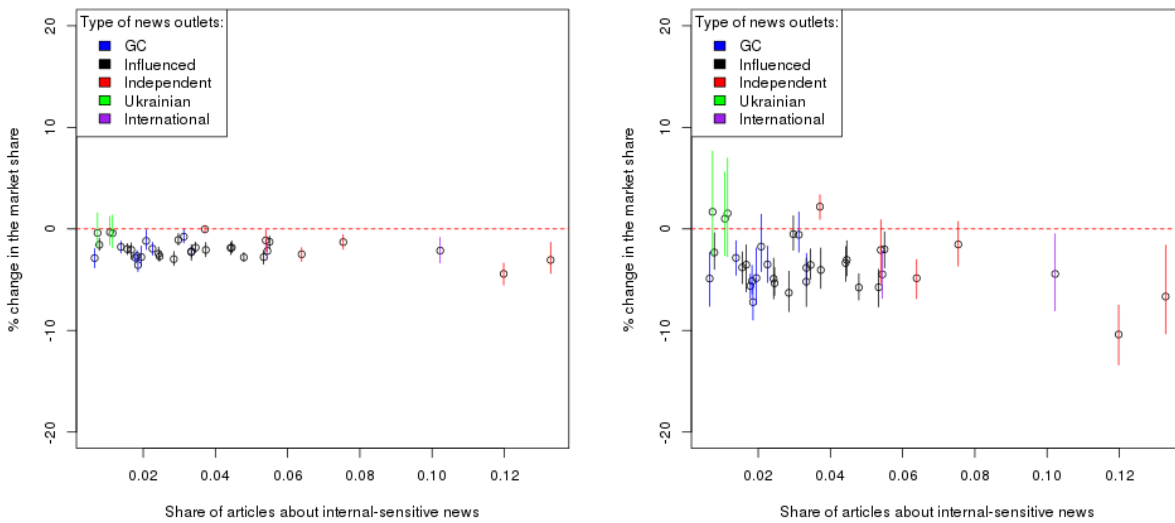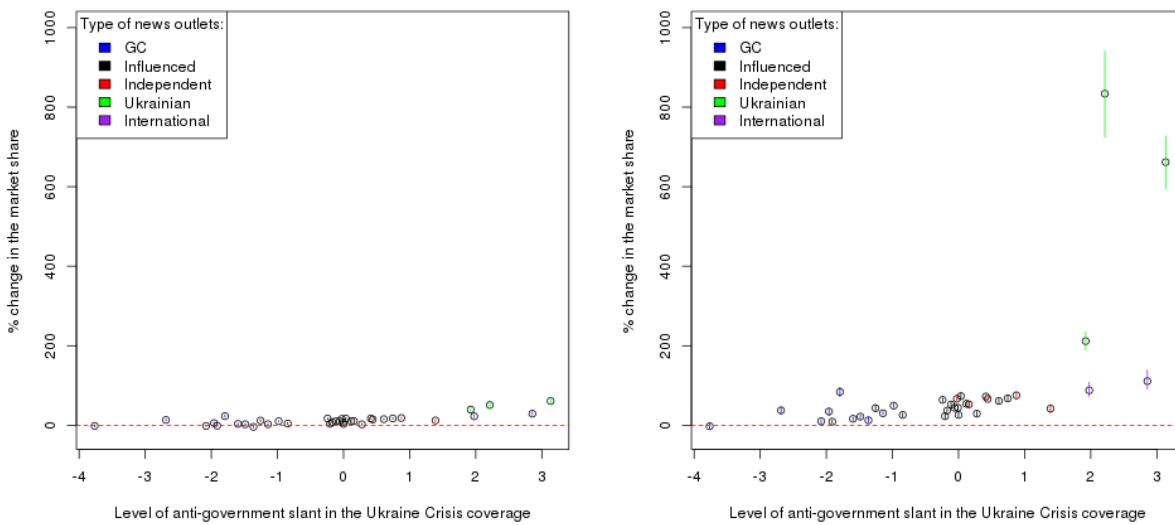

(a) $F_t^{IS} = 1$ (b) $F_t^{IS} = 3$

Dots represent the % change in the market shares of the news outlets on the days with internal-sensitive news as compared to the days with no internal-sensitive news. News outlets are sorted by the share of articles about internal-sensitive news ($\bar{F}_j^{IS}$). Solid lines represents 95% credibility interval of the estimates.

Figure 20: Change in predicted choice probabilities, $\Pr(y_i = j)$, on days with (a) an average and (b) three times the average amount of internal-sensitive news observed in the sample under homogenous preferences for slant in the Ukraine-crisis coverage, γ^- and γ^+.
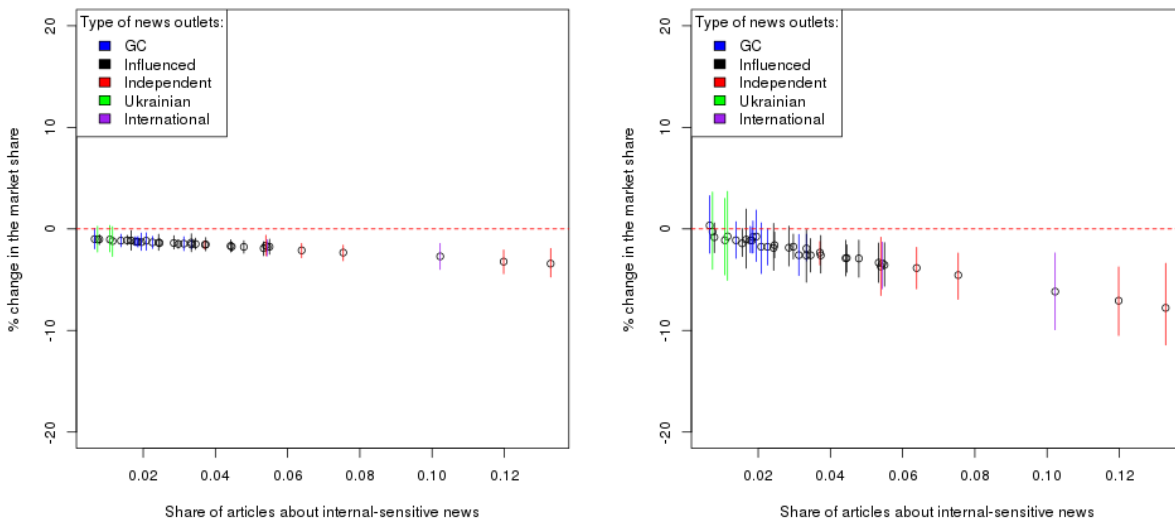

(a) $F_t^{Ukr} = 1$ (b) $F_t^{Ukr} = 3$

Dots represent the % change in the market shares of the news outlets on the days with internal-sensitive news as compared to the days with no internal-sensitive news. News outlets are sorted by the level of anti-government slant in the Ukraine news coverage (V_j^-). Solid lines represents 95% credibility interval of the estimates.

internal-sensitive news coverage, $\hat{\beta}_i^{IS} = \hat{E}_i(\beta_i^{IS})$, and permuted (across consumers) prefer-ences for internal-sensitive news topic, η_i^{IS}. Figure 21 presents the results. Once we removes sorting, independent news outlets lose more market share on the days with internal-sensitive news – on a day with 3 times the average amount of internal-sensitive news consumers are 3.67% less likely to read news from independent news outlets. Comparing this results to Figure 19, we can conclude that sorting does not play a big role in changes of the market shares, explaining only 11% of the difference in the market share between homogenous and heterogenous β^{IS} preferences cases. Similarly, we examine sorting in the case of Ukraine crisis news, and also find that it has only limited effect on the market shares.

Figure 21: Change in predicted choice probabilities, $\Pr(y_i = j)$, on days with (a) an average and (b) three times the average amount of internal-sensitive news observed in the sample under homogenous preferences for internal-sensitive coverage, β_i^{IS}, and permuted preference for internal-sensitive news, η_i^{IS}.

(a) $F_t^{IS} = 1$ (b) $F_t^{IS} = 3$

Dots represent the % change in the market shares of the news outlets on the days with internal-sensitive news as compared to the days with no internal-sensitive news. News outlets are sorted by the share of articles about internal-sensitive news ($\bar{F}_j^{IS}$). Solid lines represents 95% credibility interval of the estimates.

Alternative theory is that heterogeneity in β^{IS}, γ^- and γ^+ affects the choice probabilities directly. For example, while there is only a minority of consumers who prefer internal-sensitive news coverage, the behavior of these consumers drives an increase in the market shares of independent news outlets on the days with a lot of internal-sensitive news. To test

the direct impact of the marginal consumer preferences for internal-sensitive news coverage, β^{IS}, on choice probabilities, we permute β_i^{IS} estimates across the consumers and recompute the predicted choice probabilities. Figure 22 shows changes in the predicted choice probabilities under the permuted preferences. Results are qualitatively very similar to Figure 17: with an increase in the amount of internal-sensitive news, the expected choice probabilities decrease for a GC news outlet and increase for an independent news outlet. On a day with 3 times the average amount of internal-sensitive news an average consumer is 39.8% more likely to navigate to independent news outlets and 2.9% less likely to visit a GC news website. If we compare the magnitudes, under the permuted preferences the expected choice probabilities of the independent news outlet grow even more (39.8% compared to the 16.9% in the baseline case). Thus, correlation structure of consumer preferences decrease the degree of substitution on the days with a lot of sensitive news, and we can conclude that preference heterogeneity in β^{IS} drives increases in the market shares of the independent news outlets.[74]

Figure 23 summarizes the results. For the days with an increase in the amount of internal-sensitive news (Subfigure a), the independent news outlets gain the most. This increase is explained by the marginal distribution of β_i^{IS}, with sorting of consumers with high $\alpha_{Ind} - \bar{\alpha}$ on η_i^{IS} having no effect on the market shares of the independent news outlets. For the days with an increase in the amount of Ukraine-crisis news (Subfigure b), the Ukrainian news outlets gain the most, and changes in the market shares closely match the predicted market shares under the homogenous consumer preferences.

---

[74]Further analysis shows that the degree of substitution is reduced primarily due to the correlation between β^{IS} and persistent preferences of consumers for particular news outlets, α_j. We re-do this analysis for γ^- and γ^+ in the Ukraine-crisis news, and come to a similar conclusion.

Figure 22: Change in predicted choice probabilities, $\Pr(y_i = j)$, on days with (a) an average and (b) three times the average amount of internal-sensitive news observed in the sample under permuted preferences for internal-sensitive coverage, β_i^{IS}.
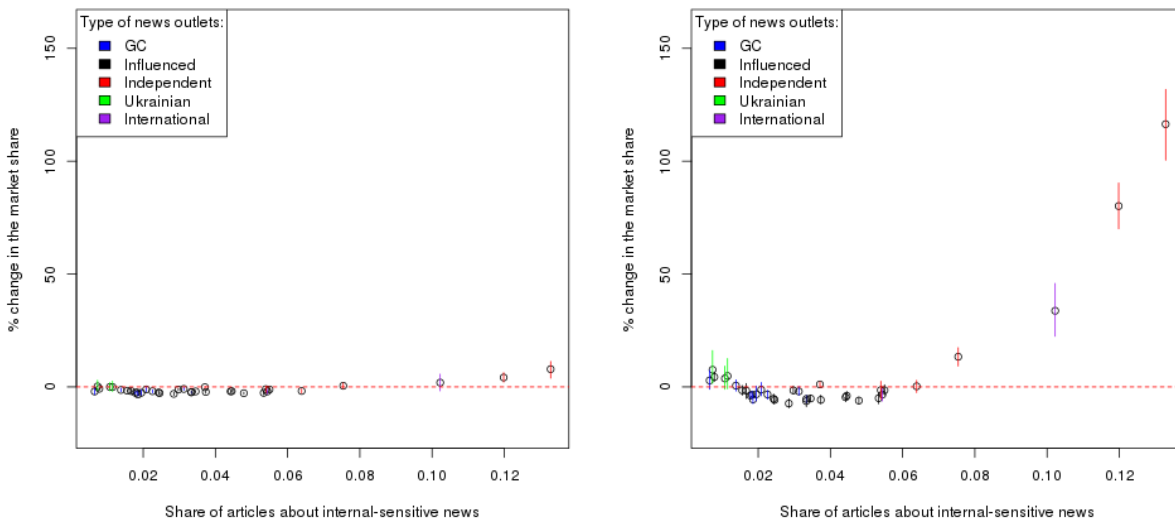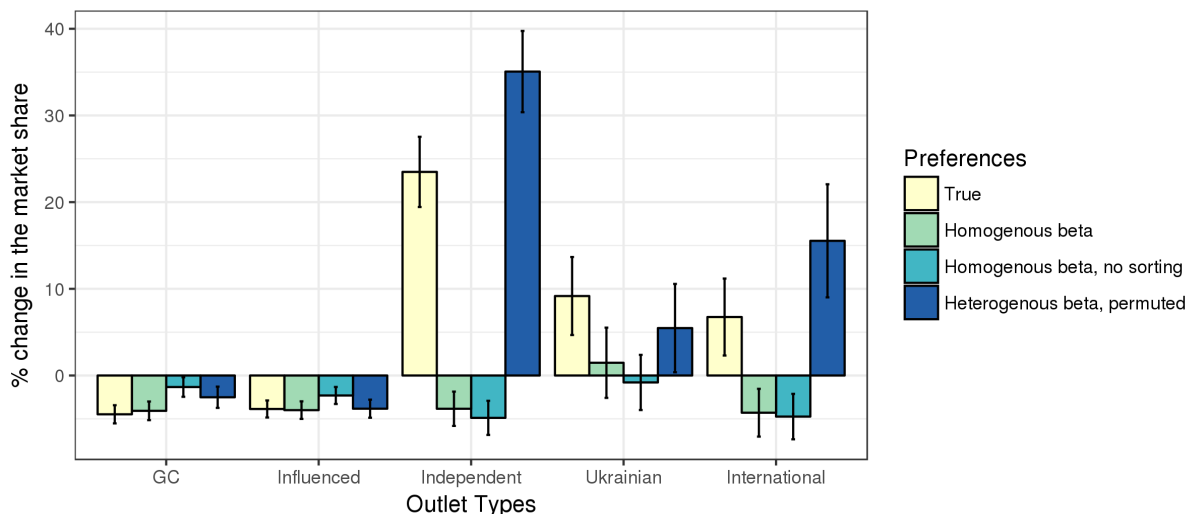

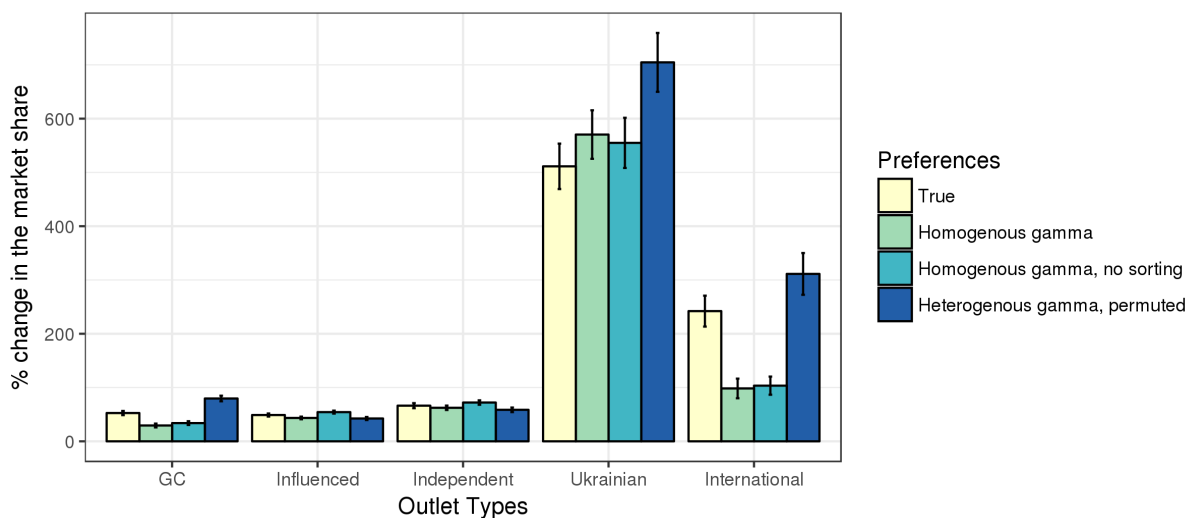(a) $F_t^{IS} = 1$ (b) $F_t^{IS} = 3$

Dots represent the % change in the market shares of the news outlets on the days with internal-sensitive news as compared to the days with no internal-sensitive news. News outlets are sorted by the share of articles about internal-sensitive news ($\bar{F}_j^{IS}$). Solid lines represents 95% credibility interval of the estimates.

Figure 23: Change in predicted choice probabilities, $\Pr(y_i = j)$, on days with three times the average amount of (a) internal-sensitive news and (b) Ukraine-crisis news observed in the sample, by a type of preference structure.

(a) Internal-sensitive news, $F_t^{IS} = 3$

(b) Ukraine-crisis news, $F_t^{Ukr} = 3$

Bars represent the % change in the market shares of an average news outlet on the days with 3 times the average amount of sensitive news, as compared to the days with no sensitive news. News outlets are grouped by types. Different preference structures correspond to (1) true estimated preferences, (2) homogenous preference for ideological content, $\beta_i^{IS} = E(\beta_i^{IS})$ (Subfigure a) and $\gamma_i^- = E(\gamma_i^-)$, $\gamma_i^+ = E(\gamma_i^+)$(Subfigure b), (3) homogenous preferences for ideological content with no sorting, η_i^{IS} (Subfigure a) and η_i^{Ukr} (Subfigure b) permuted across individuals and homogenous β_i^{IS} (Subfigure a) and γ_i^-, γ_i^+ (Subfigure b), and (4) heterogenous preferences for ideological content, uncorrelated with other preferences, β_i^{IS} (Subfigure a) and γ_i^-, γ_i^+ (Subfigure b) permuted across individuals. Solid lines represents 95% credibility interval of the estimates.

69

9.9 Correlation in Persistent Preferences

Do higher α estimates represent a higher quality of the GC news outlets, a results of 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, α_{ij}, across the news outlets. If α_{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 α_{ij} 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 24 summarizes the estimates of correlation in persistent outlet preferences, α_{ij}, across the news outlets. Similar to Table 10, we subtract the average preference for news outlets, $\bar{\alpha}_i$, from the α_{ij} to exclude the influence of consumer i's preference for news in general. News outlet are colored by their types and correspond to the legend in Figure 10 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 outlet 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_j^{IS} and F_j^{Ukr}, 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 18 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, 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 tend to have 4.56% more correlated persistent preferences of consumers. However, quality of the news outlets also plays a role, with news outlets that are 1 standard deviation more

Figure 24: Posterior estimates of the correlation matrix of persistent consumer preferences for news websites, $\alpha_{ij} - \bar{\alpha}_i$.
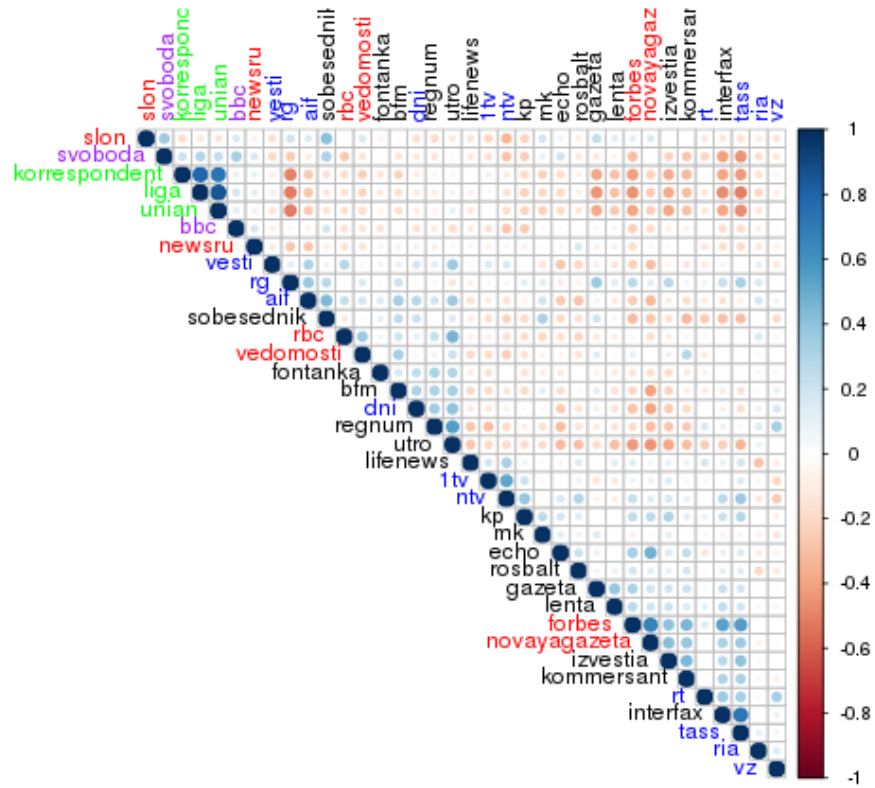
Each dot represents the correlation of $\alpha_j - \bar{\alpha}_j$ for two news outlets. Scale on the right explains the color code of the correlations. Colors of the text labels correspond to types of the news outlets used throughout the draft (first explained in Figure 10.

similar in α_j having 3.23% more correlated persistent preferences of consumers.

Table 18: Relationship between the correlations in persistent preferences of consumers, $\alpha_{ij} - \bar{\alpha}_i$, and distance between the outlets' characteristics.

	Dependent variable:
	$\mathrm{cor}(\alpha_{ij} - \bar{\alpha}_i, \alpha_{ij'} - \bar{\alpha}_i) \ \forall \ j \neq j'$
Intercept	0.1268***
	(0.0196)
$\lvert F_j^{IS} - F_{j'}^{IS} \rvert$	-0.0173*
	(0.0092)
$\lvert F_j^{Ukr} - F_{j'}^{Ukr} \rvert$	-0.0283**
	(0.0085)
$\lvert V_j^- - V_{j'}^- \rvert$	-0.0456***
	(0.0078)
$\lvert V_j^+ - V_{j'}^+ \rvert$	-0.0041
	(0.0102)
$\lvert \bar{\alpha}_j - \bar{\alpha}_{j'} \rvert$	-0.0323***
	(0.0097)
Observations	630
R^2	0.1622
Adjusted R^2	0.1555
Note:	*p<0.1; **p<0.05; ***p<0.01

Results above suggest that persistent preferences of consumers, α_{ij}, represent both the quality of the news outlet and some brand capital that news outlet has accumulated over time, with an ideology of the news outlet playing a role in the brand capital formation.