

What Drives Demand for Government-Controlled News? Evidence from Russia*

Andrey Simonov
Columbia University

Justin Rao
HomeAway, Inc.

15th February 2020

News consumers in many authoritarian countries read government-controlled sources even when independent sources are available. We test whether these choices reflect a preference for pro-government coverage versus persistent tastes for specific websites. We exploit textual data from news publications to detect government-sensitive topics and describe outlets' reporting, and detailed browsing data to trace individual-level consumption. Consumer tastes are identified from changes in consumption in response to exogenous shifts in the volume of sensitive news. Structural estimates of demand reveal that the average consumer has a distaste for pro-government ideology but strong persistent tastes for the state-owned outlets, with the latter primarily driven by third-party referrals and coverage about celebrities and sports.

*Simonov (corresponding author): asimonov@gsb.columbia.edu; Rao: justinmrao@outlook.com. We thank Charles Angelucci, Ruben Enikolopov, Andrey Fradkin, Matt Goldman, Elisabeth Honka, Sanjog Misra, Chris Nosko, Andrea Prat, Miklos Sarvary, Bradley Shapiro, Raluca Ursu, Ali Yurukoglu, participants of the 16th Trans-Atlantic Doctoral Conference, 2016 AMA-Sheth Doctoral Consortium, 2016 Economics of Media and Communications Conference, 2017 Marketing Science Conference, 15th Media Economics Workshop and 4th Media Bias Conference, Behavioral Insights from Text at Wharton, seminar participants at the University of Chicago, Stigler Center at Chicago Booth, HKUST, University of Minnesota-Carlson, University of Rochester-Simon, UCSD-Rady, Duke-Fuqua, LBS, Columbia, Southern Methodist University-Cox, UC Berkeley-Haas, Stanford, CMU, UCLA-Anderson, University of Southern California-Marshall, Moscow State University, University of Washington-Foster, NYU-Stern and the FCC for advice and suggestions. We thank the team of five independent research assistants for their help with manual text classification. Andrey is particularly grateful to his advisors and dissertation committee members Pradeep Chintagunta, Jean-Pierre Dubé, Matthew Gentzkow and Günter Hitsch for their guidance and support. All opinions represent our own and not those of our current or past employers. All remaining errors are our own.

“[...] if we don't wrap our propaganda news in the Voice show, cartoons and good TV series, nobody will watch them at all”

ANONYMOUS EMPLOYEE OF CHANNEL ONE RUSSIA
DECEMBER 19, 2019 (MEDUZA.IO, 2019)

1 Introduction

On May 2, 2014, an unprecedented outbreak of violence between the supporters and opposers of the new Ukrainian government in the city of Odessa led to 48 deaths, a story widely covered by the Russian media. However, the coverage of the event and its aftermath differed drastically across the news outlets. Independent Russian news outlets and international news outlets with Russian coverage reported that both supporters and opposers of the new Ukrainian government were throwing Molotov cocktails that could have caused the fire and that the fire had likely started due to the actions of the government opposition members who were inside the building. In contrast, government-controlled (GC) Russian news outlets reported that radical Ukraine government supporters were to blame. The coverage of the GC news outlets was characterized both by traditional media slant, or the choice of facts and language used to describe the event, and objectively false information, exemplified by the title of one of the articles overstating the number of casualties: “116 people burned alive by fascists in Odessa” (vesti.ru, 2014).

These two drastically different takes on a particular story characterize a typical choice set of the online news consumers in many authoritarian countries. Both independent news outlets, the ones that are not owned by nor influenced indirectly by the government, and news outlets that are either government-owned or influenced co-exist in this market. Almost none of the Russian news websites have a subscription firewall, so switching from one to another and finding preferred news content is simple. And still, with such availability of the independent news options, many consumers choose to read the news from the GC outlets and not from the independent outlets. In case of the Russian online news market in late 2014, 4 out of the top 5 outlets are either government-owned or potentially influenced.¹

In this paper, we aim to understand what drives this demand for the GC news outlets in authoritarian countries like Russia. We distinguish among two potential explanations.

¹This is based on public statistics (liveinternet.ru, 2014) and confirmed by our browsing data. For the news outlets classification, see Table 1.

On the one hand, consumers might read the GC outlets because of the preference for the pro-government bias in sensitive news coverage. For instance, such preferences can stem from a taste for like-minded news (Gentzkow and Shapiro, 2010) of consumers who prefer political news slanted in favor of the Russian government – as suggested by the 80% approval rating of President Putin in 2014-2016 (Economist, 2016) – or from the readers’ interest in the pro-government news framing due to “conscientious” news consumption (Mullainathan and Shleifer, 2005). To understand these tastes, one needs a revealed-preference measure – since surveys might be unreliable in countries with limited freedom of speech (Kuechler, 1998) – and an empirical strategy to identify the underlying mechanisms.

On the other hand, consumers might have a distaste for the GC news outlets’ ideological positions but have strong persistent tastes for visiting these outlets. A number of factors could contribute to these outlet-level drivers of demand – modern news websites, availability of video content, referrals by news aggregators and other third parties, accumulated brand capital of the outlet, and other sources of product differentiation that are not related to the outlet’s ideology. Throughout the paper, we refer to these factors as “persistent preferences” of consumers for news outlets. These outlet-level persistent preferences have important implications for the ability of the government to exercise media capture. If GC outlets can drive news consumption through the persistent preferences of the readers and despite the readers’ distaste for the pro-government ideological coverage, the government has an effective method of control over the ideological news diet of the readers. This would imply that the government does not need to control all news producers in the market to capture consumers’ attention (Besley and Prat, 2006); instead, it needs to invest in the quality of a handful of controlled outlets and let them compete with other news producers.

We separate out the persistent component of consumers’ preferences for the GC outlets from their preference for the pro-government slant in the sensitive news reporting by building and estimating a demand model for online news. Our identification argument relies on a novel observation that ideological positions of news outlets have a higher impact on the readers’ outlet choice on days with a higher volume of realized sensitive news. Intuitively, on days with no sensitive events to report, both the GC and independent news outlets would cover only non-sensitive news, unaffected by their ideological positions. On such days, consumers will choose a preferred news outlet based on their persistent outlet taste. In contrast, on days with a lot of government-sensitive events, the ideological positions of the GC and independent outlets would be reflected in their news reporting, and consumers would take these positions into account when making the outlet choice. Any systematic changes in

consumers’ choices would reveal their preferences for the ideological slant of the outlets and will help us identify the underlying mechanism.

The focus of this paper is the online news market in Russia in 2013-2015. In this time frame, online news consumers in Russia had a choice between a large number of established outlets owned by the government and independent from the government, as well as multiple outlets in-between – ones that were formally independent but with ties to or affected by the government. Consumers could also read versions of international and Ukrainian outlets in the Russian language. We use the information on ownership structure (Djankov et al., 2003) and reports of alleged government influence to classify the top 48 online news outlets in one of these groups. Further, we collect all accessible publication records – 3.9 million online news articles – written by these outlets between March 2013-April 2015. News articles data include the article URL, date, title and text.

We exploit the outlet classification and publications records to detect news topics that are sensitive for the Russian government. For this, we compare the publication records of the GC and other outlets, looking for the differences in news coverage that apply to *all* or *most* GC outlets – since government control should apply to all captured outlets, both in terms of which topics they censor and which ideological framing they use. This objective is different from common methods of text classification used in the literature (Gentzkow and Shapiro, 2010; Gentzkow et al., 2017) that search for language that is most *predictive* of the outlet type. We propose a simple classification algorithm tailored for our objective; the algorithm ranks outlets by the share of usage of tokens in texts – such as word unigrams – and looks for text objects for which all GC outlets systematically get high or low ranks. In the validation exercise done with manual word coding, our algorithm outperforms all feasible alternative methods of sensitive news detection.

We detect two major government-sensitive news topics in the publication records. The first topic is defined from a set of named entities that are systematically underused by the GC outlets, likely due to censorship. These events mainly correspond to political protests, opposition and corruption (hereafter “POC” news). The GC outlets systematically report only around 41.7% of POC news reported by the independent outlets, and this share is uncorrelated with the number of POC news on that day. We find almost no language differences in the POC news coverage done by GC and independent outlets – suggesting that censorship is the main method of government control of these topics.

The second government-sensitive topic is the Ukraine crisis of 2013-2015. For the Ukraine-crisis news, we show that the GC and Ukrainian outlets use systematically different language

to describe the news events, which allows us to characterize the ideological framing of the news by each outlet. For instance, the GC outlets report that Russia has “reunited” with Crimea, whereas Ukrainian outlets characterize it as an “annexation” and “occupation” of Crimea by Russia. These language differences fit well with the reports of independent journalists monitoring the news coverage of the Ukraine crisis. We use this pro-Russia and pro-Ukraine language to construct a measure of ideological framing of Ukraine-crisis news by each outlet. Our ideological framing measure closely tracks a manual classification done by two independent research assistants; the correlation in the two measures is 84%.

The sensitive news classification provides two important ingredients to our empirical strategy. First, it gives us a measure of the relative importance of sensitive news over time, which we construct as a share of articles about the sensitive news topic on a given day across all outlets. We treat this measure as an exogenous variable that is determined by the day-by-day news realizations.

Second, we characterize the sensitive news reporting and ideological framing of the news outlets. News outlets hold relatively stable reporting and ideological positions, showing a limited supply-side reaction to changes in the relative importance of sensitive news over time. We approximate the ideological positions of the news outlet by their average share of reporting about sensitive news and average share of articles with ideological framing.

The final ingredient of our empirical strategy is the individual-level news consumption data. We leverage a large panel of browsing records from Internet Explorer (IE) Toolbar data to construct individual-level records of news consumption for our sample of 48 outlets. There are 284,574 IE Toolbar users who visited these websites at least once between November 2013-April 2015. While the data suggests that IE Toolbar users are older, less interested in entertainment websites and more likely to visit business-focused news websites than an average internet user in Russia,² changes in the consumption of the top 7 online news outlets from IE Toolbar data closely track the population-level metrics (average correlation of 85.8% across outlets). Since these changes in news consumption over time are the main identifying variation that we use, we conclude that ideological preferences of IE Toolbar users are likely to match the population-level preferences. Section 3.3.1 discusses the data differences and their implications in detail.

The model-free evidence strongly suggests that the average consumer prefers more coverage of sensitive news and less pro-government ideological slant in the Ukraine crisis news. Consumers are more likely to visit Ukraine-crisis and POC than non-sensitive news articles,

²Anecdotaly this is due to a common usage of IE browser in the office setting.

particularly on independent outlets. On days with a lot of POC and Ukraine-crisis news, news outlets that report more on these topics and that have a less pro-government ideological framing get the highest increases in market shares. In contrast, GC outlets appear to benefit from traffic referred by third parties – such as the largest news aggregator, Yandex News, and other GC outlets – as well as landings on pages with video content and news articles about celebrities and sports.

We then use a structural model of demand to quantify the degree of preferences for ideological slant as opposed to news outlet. Consistent with the descriptive evidence, the average consumer of the Russian online news has a distaste for the pro-government ideological positions of the GC outlets but a high persistent preference for these outlets. The majority of consumers in the market, 58.85% and 67.2%, prefer more coverage of the POC and Ukraine-crisis news, respectively, and 54.98% of consumers prefer less pro-government framing in the Ukraine-crisis news. Since only a minority (39.9%) of consumers in the market behave like conscientious types when reading the Ukraine-crisis news, a preference for less pro-government framing suggests that independent – not GC – outlets have a more like-minded ideology to the majority of online news consumers.

However, the vast majority (87.95%) of consumers have higher persistent preferences for the GC outlets than for the independent outlets. As a result, GC outlets have a one-third market share advantage over independent outlets on days with no sensitive news. Consumers with a strong taste for GC outlets tend to navigate directly, land on the main page, and read primarily news about celebrities, sports and international affairs as opposed to news about Russian and Ukrainian politics. Correlationally, the high persistent preferences for the GC outlets appear to be driven mainly by indirect traffic – Yandex News in particular – and non-sensitive news articles, and less so by other links (often containing video content) and news articles covering the Ukraine crisis.

Finally, we use the demand estimates to conduct several counterfactual simulations to determine the importance of consumer preferences for news outlets' market shares and media power. GC outlets sacrifice 15.3% of market share due to their pro-government ideological positions, translating to as much as \$15.6 million in foregone display advertising revenues per year. In contrast, the Russian government issued \$1.21 billion in subsidies to mass media in Russia in 2015 (rbc.ru, 2015). At the same time, without the high persistent preferences of consumers, GC outlets would lose 54.3% – a 3.5 times steeper decrease compared to an effect of the inferior ideological positions. These high persistent outlet preferences of consumers also substantially increase the attention share and media power of the GC outlets

– they currently command a 33.8% attention share, but would only obtain 17.92% in the absence of persistent preferences. Once again, the data suggests that indirect traffic – such as Yandex News – and non-sensitive news articles play the most important role in forming a high attention share of customers for GC outlets. Following Prat (2017), the attention share of 33.8% could enable the government to swing 25-75% elections through media persuasion.

To our knowledge, this is the first paper to estimate a demand model for online news that separates out consumers’ ideological preferences from their (heterogeneous) persistent tastes for news outlets. We are able to estimate this model by leveraging a new identification strategy that builds on exogenous shifts in the volume of sensitive news over time, adding to other empirical strategies of estimating consumer preferences for the ideological slant (Gentzkow and Shapiro, 2010; Martin and Yurukoglu, 2017). Our model of the online news demand builds on Strömberg (2004) and Gentzkow and Shapiro (2015) and contributes to the growing empirical literature on online news markets (Gentzkow et al., 2011; Sen and Yildirim, 2016; Athey et al., 2017; Cagé et al., 2019). Our findings demonstrate the ability of governments to exercise media capture in formally free news markets and contribute to the empirical literature on the effect of government news control on consumers (Durante and Knight, 2012; Enikolopov et al., 2011; Bai et al., 2015; Roberts, 2014; Garcia-Arenas, 2016; Knight and Tribin, 2016), and inform the theoretical literature on media capture (Besley and Prat, 2006; Petrova, 2008; Prat and Strömberg, 2013; Edmond, 2013; Gehlbach and Sonin, 2014), media power (Prat, 2017), and news demand more broadly (Mullainathan and Shleifer, 2005; Xiang and Sarvary, 2007; Zhu and Dukes, 2015). We describe the role of alternative mechanisms behind the persistent preferences of consumers, including third-party referrals, non-sensitive news articles, other content and inertia in consumer choices. Our text classification algorithm also contributes to the literature measuring media censorship and framing (Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010; Gentzkow et al., 2019) by proposing a simple new method that outperforms existing methods in our context. Finally, our analysis of positioning of GC outlets complements other work on media control and product differentiation in autocracies (Qin et al., 2017, 2018).

The next section builds a stylized model of demand for news and lays out our identification strategy. Section 3 describes the Russian online news market and our data sources. Section 4 describes the classification of the government-sensitive news and characterizes the reporting of the news outlets. Section 5 presents model-free evidence on consumer preferences for the GC outlets. We describe our empirical specification in Section 6 and present the demand estimates and counterfactual simulations in Section 7. Section 8 concludes.

2 A Stylized Model and Identification

In this section, we present a stylized model of the news supply and demand in the markets with partial government control and lay out our identification strategy.

2.1 Basic Model

Suppose there are two news outlets in the market, A and B. Every day, these news outlets produce one unit of news product, such as a newspaper or a set of articles on a website.

The news product consists of commodities of two types: news articles that are sensitive and those that are not sensitive for the incumbent government. For now, assume that any publications about sensitive events are bad for the government; the government is indifferent about the non-sensitive news publications.

Consumers have stable and heterogeneous preferences for sensitive and non-sensitive news articles. Assume that at day t consumers choose at most one outlet or decide not to consume the news altogether. Consumer i chooses an option with the highest utility among

$$U_{ijt} = \beta_i x_{jt}^S + \lambda_i x_{jt}^{NS} + \epsilon_{ijt} : j \in \{A, B\}, \{x_{jt}^S, x_{jt}^{NS}\} \in [0, 1], \quad (1)$$

where x_{jt}^S and x_{jt}^{NS} are the amount of sensitive and non-sensitive news in the outlets j 's coverage, respectively, and ϵ_{ijt} is an unobserved idiosyncratic shock to the consumer's utility. Following the standard discrete-choice literature (Train, 2009), we can derive consumer demand for news outlets' products $\{D_A, D_B\}$, which is driven by the distribution of consumer preferences, $\{\beta, \lambda\}$, commodity choices of the news outlet, $\{x_{jt}^S, x_{jt}^{NS}\}$, and the distribution of the idiosyncratic shocks, ϵ_{ijt} .

News outlets make daily production decisions on the amount of sensitive and non-sensitive news commodities in their product, x_{jt}^S and x_{jt}^{NS} . The news commodities are costly to produce as they require journalists to investigate the news topics. However, it is less costly to produce news about a certain topic on the days when a lot of topic-related events happen. For example, writing sensitive news is more costly on the days when there are no sensitive news events as production requires more investigation. More formally, news production costs $c_t^S(x_{jt}^S, V_t^S)$ and $c_t^{NS}(x_{jt}^{NS}, V_t^{NS})$ are decreasing in the amount of the events of the same type that happen on day t , $\{V_t^S, V_t^{NS}\} \in [0, 1]$.

Finally, suppose that the news outlet A is controlled by the government and the news outlet B is independent. Given that the government dislikes sensitive news publications,

it exercises censorship by imposing additional costs of production of sensitive news on the outlet A, $c^G(x_{At}^S)$.³ The shape of the $c^G(\cdot)$ function is determined by the objective function of the government.

Two observations follow from this setting.⁴ First, the controlled outlet A would choose to produce less sensitive news than the independent outlet B, $x_{At}^{S*} \leq x_{Bt}^{S*}$, as it faces higher marginal costs of sensitive news production.

Second, unless the shape of $c^G(\cdot)$ function is highly concave – meaning that the government mainly cares about the first few sensitive stories reported by the outlet A – the difference in the amount of sensitive news produced by outlets, $x_{Bt}^{S*} - x_{At}^{S*}$, is increasing in V_t^S . Intuitively, when there is no sensitive news to report, $V_t^S = 0$, it can be very costly for both news outlets to produce sensitive news (high c_t^S), so both outlets produce very low x_{jt}^{S*} . In contrast, when there is a lot of sensitive news to report, the cost of sensitive news production is low and c^G plays a more important role. In Section 4.2, we confirm that the difference in the sensitive news reporting between the GC and independent outlets increases with V_t^S . We further show that news outlets tend to report a certain share of sensitive news that does not depend on V_t^S , meaning that we can decompose $x_{jt}^S \approx V_t^S \bar{x}_j^S$, where \bar{x}_j^S is the share of sensitive news reported by the news outlet j .⁵

Our key identification argument relies on the two observations above. We use changes in the sensitive news reporting induced by V_t^S to identify consumer preferences for sensitive news, β . Before a further discussion of the identification, we extend and adjust the basic model to account for other important features of the online news consumption and production.

2.2 Extensions

Persistent preferences. Apart from the news commodities supplied, outlets can differentiate themselves in a variety of ways (Strömberg, 2004), such as website design, overall quality of the news coverage, other content of the website, and promotion by third parties.

³For example, a government that instructs a news outlet not to cover a story or omit some facts from a story about a corruption scheme organized by some officials is censorship. Media economics literature refers to censorship as “issue and fact bias” (Prat and Strömberg, 2013) or as “filtering or selection of news” (Gentzkow et al., 2016). Censorship works through the effects of agenda setting (McCombs and Shaw, 1972) and priming (Iyengar and Kinder, 1987).

⁴Online Appendix A presents an extended discussion of the news outlets’ optimization problem.

⁵Since our focus is on estimating consumer preferences, we stop short of estimating the shape of the cost functions. Instead, we take the editorial strategies of the news outlets as given, at least in the short-run.

Consumers can like or dislike these attributes of the outlets,

$$U_{ijt} = \alpha_{ij} + \beta_i x_{jt}^S + \lambda_i x_{jt}^{NS} + \epsilon_{ijt} : j \in \{A, B\}, \{x_{jt}^S, x_{jt}^{NS}\} \in [0, 1], \quad (2)$$

where α_{ij} represent the matching value between consumer i 's preferences and features of the news outlet j . These persistent preferences might also include the effects of habit formation and inertia in news consumption.

Space constraints. Up to this point we have assumed that news outlets make two separate choices of x_{jt}^S and x_{jt}^{NS} that only depend on the realizations of V_t^S and V_t^{NS} . In practice, outlets operate under capacity constraints; their coverage cannot exceed a certain number of articles, for example, because of a fixed amount of space in the newspaper or a limited amount of journalists and editors in the online outlet. We simplify the model by assuming that the news outlets always have to fill a strict amount of space, $x_{jt}^S + x_{jt}^{NS} = 1$, so the only thing that varies over time is the ratio of the produced sensitive and non-sensitive news commodities.⁶

Using this simplification, we can re-write consumer utilities as

$$U_{ijt} = \alpha_{ij} + \beta_i x_{jt}^S + \lambda_i(1 - x_{jt}^S) + \epsilon_{ijt} = (\alpha_{ij} + \lambda_i) + (\beta_i - \lambda_i)x_{jt}^S + \epsilon_{ijt}, \quad (3)$$

where $\alpha_{ij} + \lambda_i$ is the persistent preference of the consumer i for a news outlet j only with non-sensitive news, and $\beta_i - \lambda_i$ is the relative preference of the consumer i for sensitive news over non-sensitive news.⁷ With a slight abuse of notation, we redefine consumer utility to get rid of λ_i ,

$$U_{ijt} = \alpha_{ij} + \beta_i x_{jt}^S + \epsilon_{ijt}, \quad (4)$$

where α_{ij} is the persistent preference of the consumer i for a news outlet j only with non-sensitive news, and β_i is the relative preference of the consumer i for sensitive news over non-sensitive news.

Ideological framing. So far, we have assumed that the only method of government control over sensitive news reporting is censorship. Apart from censorship, governments can frame the sensitive news reporting (Prat and Strömberg, 2013), making it more aligned with the government's ideology. This implies that the sensitive news reporting can have an

⁶This assumption is useful in our empirical specification of the model since we observe only the relative importance of sensitive and non-sensitive news over time, V_t^S and V_t^{NS} .

⁷Note that λ_{ij} can include any persistent difference in the non-sensitive news reporting between outlets A and B, capturing their differentiation in the non-sensitive news reporting.

ideological stand bias, such as supporting, opposing, or being neutral about the government.⁸

We extend the model and allow news outlets to choose the ideological framing in their sensitive news reporting, $sl_j \in [-1, 1]$. For instance, some outlets might choose to report that Russia has “reunited” with Crimea after the “referendum” in 2014, while other news outlets might refer to it as an “occupation” or “annexation” of Crimea after a “pseudo-referendum.”

Consumers hold stable preferences for ideological framing of the sensitive news,

$$U_{ijt} = \alpha_{ij} + (\beta_i + \gamma_i sl_j)x_{jt}^S + \epsilon_{ijt}, \quad (5)$$

where γ_i captures consumer’s preference for the ideology of the reporting – for instance, driven by their taste for like-minded news (Gentzkow and Shapiro, 2010).

Conscientious news consumption. Consumers’ preferences for the ideological framing in the news coverage might also be driven by “conscientious” news consumption (Mullainathan and Shleifer, 2005; Xiang and Sarvary, 2007). Conscientious news consumers sample alternative ideological positions to filter out the ideological framing in the news reporting. This predicts that conscientious consumers will read more ideologically-diverse news outlets on days with a lot of sensitive news coverage, whereas consumers with a preference for like-minded news will read more similar news outlets when a lot of sensitive news is covered.

We capture this difference in behavior of the conscientious consumers by borrowing from the literature on variety-seeking behavior in product choice (McAlister and Pessemier, 1982; Kim et al., 2002). A consumer gets a utility of

$$U_{\tau ij} = \begin{cases} \alpha_{ij} + (\beta_i + \gamma_i sl_j)x_{jt}^S + \epsilon_{\tau ij} & \text{if } \tau = 1, \\ \alpha_{ij} + (\beta_i + \gamma_i sl_j + \rho_i |sl_j - s_{i\tau}|)x_{jt}^S + \eta_i |sl_j - s_{i\tau}| + \epsilon_{\tau ij} & \text{if } \tau > 1, \end{cases} \quad (6)$$

where τ is the choice occasion of consumer i on day t , and $s_{i\tau}$ is the ideological framing of the outlet that was consumed on $\tau - 1$. A positive coefficient ρ_i signals an increase in the ideological “variety-seeking” of consumer i on days with a lot of sensitive news, consistent with the conscientious news consumption. In contrast, a negative ρ_i means that consumers read more ideologically-similar news outlets on days with a lot of sensitive news coverage, consistent with the like-minded news consumption.⁹

⁸The literature refers to this ideological bias as ‘framing and ideological stand bias’ (Prat and Strömberg, 2013) and “distortion of news” (Gentzkow et al., 2016).

⁹We note that this stylized model ignores any forward-looking behavior the consumer might have when choosing whether to read another article within a day. We also refrain from incorporating and testing the potential complementarities across the news outlets into the demand (Gentzkow, 2007) and supply (Xiang

2.3 Identification

Our identification strategy of consumer preferences, $\{\alpha_{ij}, \beta_i, \gamma_i, \rho_i\}$, relies on exogenous shifts in the amount of sensitive news that happens over time. Such shifts influence the volume of sensitive news reporting of the news outlets, x_{jt}^S , and thus change the importance of the ideological positions of the news outlets. The distribution of the persistent preferences of consumers, α_j , is identified from the outlet choices when there is no sensitive news to cover. Relative preferences for the sensitive news, β_i , and ideological framing, γ_i , are identified from shifts in the sensitive news reporting and ideological positions of the news outlets. The distribution of ρ , an ideological variety-seeking preference of consumers, is identified from shifts in the sensitive news reporting and ideological distance between two subsequently consumed outlets.

To estimate consumer preferences, we need measures of individual-level news consumption, changes in relative importance of sensitive news over time, and ideological positions of the news outlets. We get these measures from two separate datasets, consumer browsing histories and publication records, which we describe in section 3. We detect sensitive news and recover the ideological positions of the news outlets in section 4.

3 Data

3.1 Online News Market Structure in Russia in 2013-2015

Despite high government control over the offline news market starting in 2000, online news outlets in Russia enjoyed relative freedom up until 2013. A large number of independent players existed in the online news media landscape, the second most important source of news in Russia after the TV.¹⁰ Since the beginning of 2013, political pressure has forced a number of top online news outlets to change their editorial and management teams, including prominent cases like changes of editor-in-chief at RIA Novosti, a major state news agency with balanced news coverage, and lenta.ru, one of the largest independent news outlets.¹¹ Government control further intensified in February of 2014 with the beginning of the Ukrainian crisis – the government reacted by blocking the websites of some opposition lead-

and Sarvary, 2007) models.

¹⁰In 2014, 23% named internet as their main news source, compared to 60% who have named TV news as the main source. By 2017, the importance of internet has increased to 32% and the importance of TV news dropped to 52% (VTsIOM, 2017).

¹¹Appendix B.1 list the changes and the corresponding outlets.

ers in March 2014 (bbc.com, 2014) and implementing a law to limit the foreign ownership of Russian news outlets to 20% (squirepattonboggs.com, 2014).

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

GC	Potentially Influenced	Independent	International	Ukrainian
vz (5.17%)	lenta (6.48%)	rbc (15.3%)	bbc (1.63%)	korrespondent (1.97%)
tass (5.15%)	regnum (6.4%)	newsru (1.67%)	svoboda (0.77%)	unian (1.73%)
vesti (4.24%)	gazeta (3.66%)	tvrain (1.47%)	reuters (0.01%)	liga (0.78%)
rg (4.22%)	utro (2.83%)	vedomosti (0.8%)	meduza (0.00%)	
ntv (3.41%)	interfax (2.38%)	slon (0.75%)	dw (0.00%)	
aif (3.06%)	kommersant (2.38%)	novayagazeta (0.74%)		
ria (2.52%)	kp (2.32%)	forbes (0.68%)		
dni (1.9%)	mk (1.93%)	snob (0.59%)		
rt (1.5%)	fontanka (1.91%)	the-village (0.24%)		
1tv (0.66%)	lifenevs (1.86%)	newtimes (0.10%)		
	rosbalt (1.49%)			
	echo (1.46%)			
	izvestia (0.94%)			
	bfm (0.91%)			
	sobesednik (0.81%)			
	polit (0.40%)			
	znak (0.27%)			
	ng (0.26%)			
	ridus (0.15%)			
	trud (0.12%)			

We simplify the domain names; for instance, 1tv stands for www.1tv.ru. Most domains have the www.*.ru structure, with some exceptions. Outlet-to-type classification is done based on the media ownership information and evidence of the indirect influence listed in the Online Appendix B. We present outlet market shares computed based on news article visits in IE Toolbar data in parentheses.

A large number of online news outlets in Russia have remained active and independent by the end of our data period, April 2015. At the same time, an intensified government control has increased a number of outlets that are formally independent but might be influenced by the Kremlin. We label these outlets as “potentially influenced” – those that are not owned by the government but can face some political pressure indirectly – for instance, by the government’s pressure on the news outlets’ owners.

Table 1 presents the top 48 Russian-language news outlets.¹² We group news outlets by the degree of the (potential) government influence, determined by the ownership structure

¹²We have tried to include all significant news outlets, so the set contains even the outlets with little popularity in Russia, such as the Russian version of Deutsche Welle, www.dw.com/ru.

(Djankov et al., 2003) and evidence of the indirect influence.¹³ The first column contains outlets that are owned by the government or members of the incumbent political party, which we classify as being directly controlled by the government. The second column includes the “potentially influenced” outlets, ones that are formally independent but can be indirectly influenced by the government. Given the ambiguous degree of control over the “potentially influenced” outlets, we exclude them from the sensitive news classification. The third column of Table 1 contains independent outlets, the ones with no indication that they could be under an indirect government control. Most of these news outlets are owned either by journalists, international media companies or the government opposition. Columns four and five present the outlets with Russian language news coverage that are international, separating out the Ukrainian outlets.

3.2 Publication Records

We collected publications records of the 48 outlets described above for April 1, 2013 – 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 news articles. For each article, we collect the title, text, URL link, and timestamp.¹⁴ We process texts using standard techniques such as stemming and removing the stop words. Online Appendix C provides details about the data collection and processing.

Table 2: Descriptive statistics of news articles published by outlets.

Type	# Articles	Share of Articles (%)			# Articles/Day	# Words/Article Average outlet
		Mean	S.D.	CoV		
GC	1,168,569	29.94	2.39	0.08	161	205
Independent	449,094	11.51	1.09	0.09	80	179
Pot. Influenced	1,848,556	47.37	1.2	0.03	134	156
International	120,589	3.09	0.48	0.15	42	213
Ukrainian	315,927	8.09	1.45	0.18	151	155
Total	3,902,735					

The standard deviation of article shares is computed for each outlet type across weeks. The coefficient of variation (CoV) is the standard deviation of the article shares divided by the mean.

¹³Examples of indirect influence include the removal of news articles and firing journalists under political pressure. Online Appendix B presents more detailed information on the ownership structure and evidence of the indirect influence for each news outlet.

¹⁴For five news outlets (“meduza,” “newtimes,” “ridus,” “snob,” “the-village”), article texts were not collected for technical reasons. While we use these outlets for the sensitive news detection (exploiting titles instead of article texts), we drop them from the descriptive analysis and demand estimation due to an unreliable measure of slant estimate.

The first five columns of Table 2 present the number and share of news articles for each type. Twenty potentially-influenced outlets publish almost half (47.37%) of all the news in the sample, GC and independent outlets publishing the other 30% and 11%, respectively. International and Ukrainian news outlets publish the remaining 3.09% and 8.09% of the articles. These shares are relatively stable over time – the standard deviation of the shares of articles (computed across weeks for each type) is between 0.5 and 2.4 percentage points, and the implied coefficients of variation are between 0.03 and 0.09 for Russian news outlets and a slightly higher 0.15-0.18 for the international and Ukrainian outlets.¹⁵

The last two columns of Table 2 describe the relative size of news coverage done by different types of outlets. The news coverage of an average GC outlet is more extensive compared to the coverage of an average independent outlet; the GC outlets publish more news articles on an average day (161 versus 80) and have more words per article (205 versus 179). We further investigate the differences in the news topic coverage and ideological framing of the outlets in Section 4.1.

3.3 News Consumption Records

We measure news consumption with the browsing data from the Internet Explorer (IE) Toolbar, which includes complete browsing histories for a subset of IE users. The users included in the IE Toolbar data have installed a plug-in on their IE and opted-in for the data collection.¹⁶ IE Toolbar data contain information about each webpage consumers visited (URL), websites where consumers came from (referral URL), timestamp of the visit, number of seconds spent, browsing session ID, user ID, language of the browser, country of the user, and other information. We focus the analysis on Toolbar users who specified Russian as the language of their browser.¹⁷

Although IE Toolbar data were collected for several years, the unique user IDs were 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 collected the browsing data between November 15, 2013, and March 31, 2015, 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. While this is only 13% of users

¹⁵Figure A1 in Online Appendix D plots the article shares by types of the outlets.

¹⁶Around 75% of users who installed the plug-in opt-in to the data collection.

¹⁷Having a browser in the Russian language indicates that the user knows Russian and is potentially in the market for Russian online news.

with the IE browser set to Russian language, they account for 77.8% of all browsing. In total, our sample contains 20.27 million URL visits of the 48 news-outlet websites defined above.¹⁸ Thus, for each consumer, we observe the history of the news outlet visits on the IE browser.¹⁹ These observations include four types of web pages visited by the users: news outlets’ main pages, subdirectories, news articles, and other pages such as special projects and videos. We use URL structure and the publication records data described in Section 3.2 to classify URLs into these four groups.²⁰

Table 3: Summary of browsing behavior

Page Type	Page views		Time Spent (%)	Median Time Spent (seconds)
	(#)	(%)		
Main page	4,231,761	20.9	14.3	41
News articles	9,620,141	47.5	52.9	89
News subdirectories	2,637,716	13	14.4	63
Other	3,780,583	18.7	18.5	51
All	20,270,255	100	100	67

Table 3 shows summary statistics of browsing by types of the URLs. News articles account for around a half (47.5%) of URL visits, and more than half (52.9%) of the overall time spent on these webpages. A median visit to a news article URL takes 89 seconds. The main page accounts for 20.9% of all visits, and news subdirectories and other pages account for 13% and 18.7%, respectively.

Table 4 summarizes main sources of the referral traffic for the news outlets. In more than a majority (53.6%) of the first website visits on a day, consumers navigate to the website directly (there is no referral recorded), with Yandex being the second most common traffic source, accounting for 21.7% of the visits.²¹ Other browsers, such as Google, Bing and Rambler, account for 7.5% of the first visits, other news aggregators apart from Yandex – 2.1%, and 1.3% of the first visits is referred by other news outlets in our sample. Social media traffic accounts only for 0.34% of website landings in our sample – reflecting a low

¹⁸There are 26.54 million page views in the data. We combine multiple subsequent page views of the same URL by the same user to one URL visit. Such subsequent page views occur if consumer makes a click on the page without changing page URL – for instance, while scrolling through page photos.

¹⁹We observe only news consumption of the users from the same browser – a consumer might have more online news consumption occasions on the same day. Given that we do not have access to the cross-device data, we have to assume that the user has the same reading patterns across the devices and browsers.

²⁰Online Appendix G contains details of this classification.

²¹Most of this traffic is coming from `news.yandex.ru`, a popular news aggregator run by Yandex.

role of social media in the online news market in Russia at that time.²² If we zoom in only on the news articles, direct navigation and clicks on own website correspond to 68.6% of all visits, and Yandex accounts for 16.1%.

Table 4: Summary of referral websites

Referral from:	First Time in a Day (%)	News Articles (%)	All URLs (%)
Direct & from This News Outlet	53.62	68.63	77.84
Yandex	21.70	16.09	9.76
Other Browsers (not Yandex)	7.51	4.02	3.64
Other Aggregators (not Yandex)	2.11	1.90	1.15
Other News Outlets	1.31	0.86	0.65
Social Media	0.34	0.24	0.16
Other Websites	13.42	8.26	6.80

Column “first time in a day” corresponds to a first news outlet visit of a consumer on a given day.

3.3.1 IE Toolbar Representativeness

Before we proceed with the analysis, we examine whether news consumers in the IE Toolbar data are representative of the overall population of news consumers in Russia. While the market share of the IE browser in Russia in November 2013-March 2015 was a sizable 14.4%, following Chrome (42.9%) and Firefox (18.7%) browsers (statcounter.com, 2015), we are concerned that there is a systematic difference in news and ideological preferences between the IE users and general population.

To make this comparison, we collected population-level data on daily visits of the most popular websites in Russia using `liveinternet.ru` (LI), a website that tracks statistics for the Russian internet. Due to the layout of the website ranking on LI, we can collect reliable records of usage over the period of time studied for the 30 most popular websites in Russia, which includes seven news websites from our sample.²³

Online Appendix H compares browsing habits of the news consumers in the IE Toolbar data to the general population. Results suggest that IE Toolbar users are older, less interested in streaming and entertainment websites and more interested in news than the general population tracked by LI. This is consistent with anecdotes that the IE browser is more

²²This is contrast to the US and Europe, where news consumption through social media (e.g. Facebook) is more common.

²³We use the digital archive “Wayback Machine” to collect historical data on website usage. The top page includes only the top 30 websites; Wayback Machine does not have frequent records for the other pages.

likely to be used by office workers. At the same time, the overall rankings of the websites are relatively similar, with the same top 5 websites in both IE and LI datasets, and five out of the top seven news outlets in the LI data also present in the top seven in the IE Toolbar data. The main difference in news outlets' visits in IE and LI datasets stems from a higher market share of rbc.ru, a business-focused news agency, and a lower market share of ria.ru, a news agency competing with rbc.ru, in the IE data. Once again, this is consistent with the anecdotes that IE users are more business-focused.

We further compare news consumption in the IE and LI data by looking at changes in news website visits over time. This step is particularly important since temporal variation in news consumption is the key identifying variation in our model. Figure 1 presents the normalized average traffic to the top seven LI news outlets based on the LI and IE Toolbar data. Changes in the news consumption in the IE Toolbar data closely track the population-level consumption in the LI data, with a correlation of 0.858. Figure A10 in Online Appendix H presents changes in the traffic for each of the top seven news outlets. The correlations between traffic changes in the LI and IE Toolbar datasets vary from 0.52 to 0.914. In particular, correlations for rbc.ru and ria.ru are 0.914 and 0.702, respectively, showing representativeness of changes in news consumption in IE Toolbar data even for these over- and under-sampled websites.

We conclude that while IE Toolbar data oversamples business-oriented news readers compared to the population of news consumers in Russia, consumption habits of the IE Toolbar users are otherwise representative of the news consumption of the population.

4 Government-Sensitive News

In this section we use publication records data to detect and describe government-sensitive news topics, the volume of sensitive events happening over time, and news outlets' reporting on these topics.

4.1 Detection of Government-Sensitive News

The key product differentiation decision of news outlets is which news topics to cover and how to present them. For news outlets under government control, this differentiation decision is influenced by the sensitivity of news topics and the media control strategy of the government. We start with describing main dimensions of news outlets' differentiation in the market and detecting news that are reported differently by the GC outlets.

Figure 1: Normalized average number of weekly visitors to the top seven news outlets, IE Toolbar and LI data



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

News coverage is represented by the publication records data. Following the literature on event detection (Allan et al., 1998) and topic modeling (Blei et al., 2003), we treat article texts as collections of words or n-grams, a “bag-of-words” approach. Such collections of words can be indicative of news topics and ideological framing of the news stories.

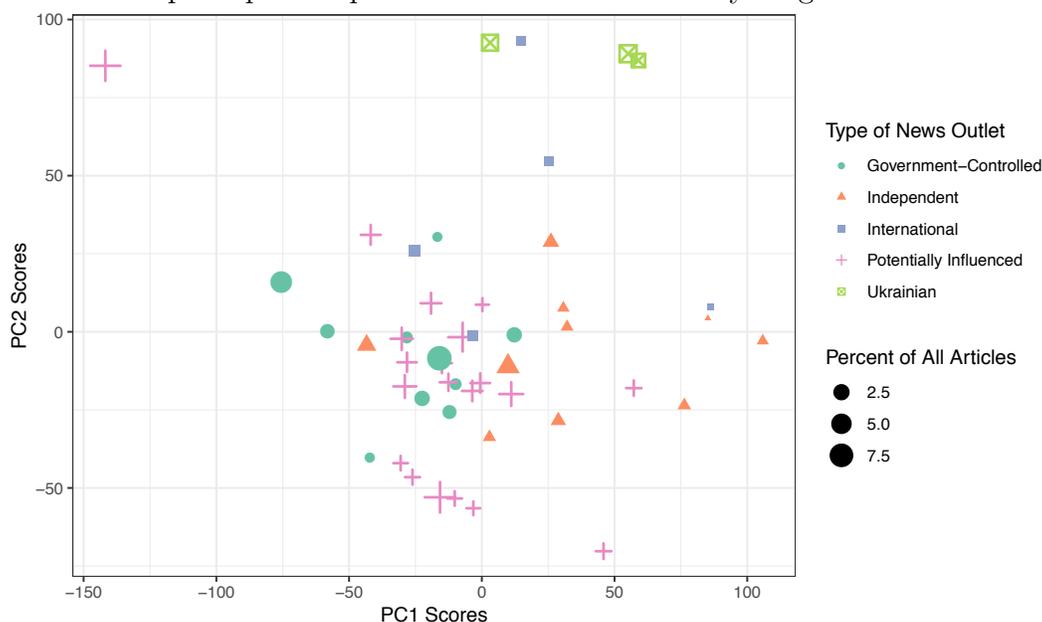
We define news topics using the universe of named entities present in the article texts. Named entities correspond to the information about actors (people or organizations), locations and timing of the news events, which are crucial in describing the news events (Hu et al., 2013).²⁴ Tracking named entities is a common approach in the information retrieval literature to extract news representations (Kumaran and Allan, 2004, 2005); named entities can successfully define news topics (Kim et al., 2012) and increase news topic coherence when getting more weight in the topic model (Krasnashchok and Jouili, 2018). We use a simple named-entity recognition system that searches for capitalized names in texts, detecting 21,873 unigrams and 16,917 bigrams of named entities in the texts that appear more than 200 times.²⁵

²⁴For example, a title of one of the top news stories on the day when this paragraph was written, “Panama Paper: David Cameron’s worst week as Prime Minister,” contains named entities (proper nouns) “Panama Papers,” “David Cameron,” and “Prime Minister,” which summarize the topic of the news article but do not capture the sentiment of this topic (captured by the word “worst”).

²⁵Online Appendix C provides more details on named-entity detection. We keep only relatively common words to make sure that they refer to an important topic. The threshold of 200 times is chosen arbitrarily. Local changes in the threshold does not affect the results.

A collection of topics covered by each news article or outlet is represented by a (long) vector of counts of the named entities that appear in the texts. We examine the main dimension of differentiation in topic coverage across the news outlets by extracting principal components (Qin et al., 2018) from a 48 by 21,873 matrix of normalized named entities counts.²⁶

Figure 2: First two principal component scores of named entity usage across the news outlets.



Color of the dots represent the type of the outlet and size represents the share of articles published by this outlet.

Figure 2 summarizes scores of the first two principal components across the outlets in a product differentiation “map.” The scores are colored and sized based on the type and size of the outlet they represent, respectively. The first principal component almost perfectly separates out GC and independent outlets – 16 out of 24 outlets with the scores above the median are either independent, international or Ukrainian, and only one is GC. In contrast, 9 out of 24 outlets below the median score are GC and 13 are potentially influenced. The second principal component differentiates outlets on the volume of coverage of news about the events in the Ukraine, which is evident from the Ukrainian outlets having the top scores. Interestingly, the GC outlets are clustered closely together, suggesting that the product differentiation among them is limited.

²⁶Counts are normalized by the overall usage of named entities by a news outlet to correct for differences in outlets’ size.

We conclude that the difference in coverage between independent and GC news outlets is the main differentiation dimension in the online news in Russia, likely driven by the difference in sensitive news coverage. This is partially supported by the nature of named entities with the highest loadings in the first principal component. The top-20 words include “Navalny”, the last name of a prominent opposition leader, and “Roskomnadzor”, a censorship agency in Russia. At the same time, journalist names (“Kashin”, “Venedictov”) and words corresponding to more general topics (“Wikipedia”, “Putin”, “Yandex”, “Spotify”) are also present in the list, meaning that pooling words together when running the principal component analysis can combine informative and incidental words. We separate out informative and incidental words by running a classification algorithm that detects systematic differences in coverage on the word level; we describe this classification next.

4.1.1 Censored News: Political Protests, Opposition and Corruption

To recover a set of topics sensitive for the government, we look for unigrams and bigrams of named entities that are systematically overused or underused by the GC outlets. We run this search separately for each named entity to avoid pooling informative and incidental words, like in the case of the PCA analysis.²⁷ Further, a difference in usage should not be driven by one or two outlier outlets. To capture these ideas, we propose the following simple classification algorithm:

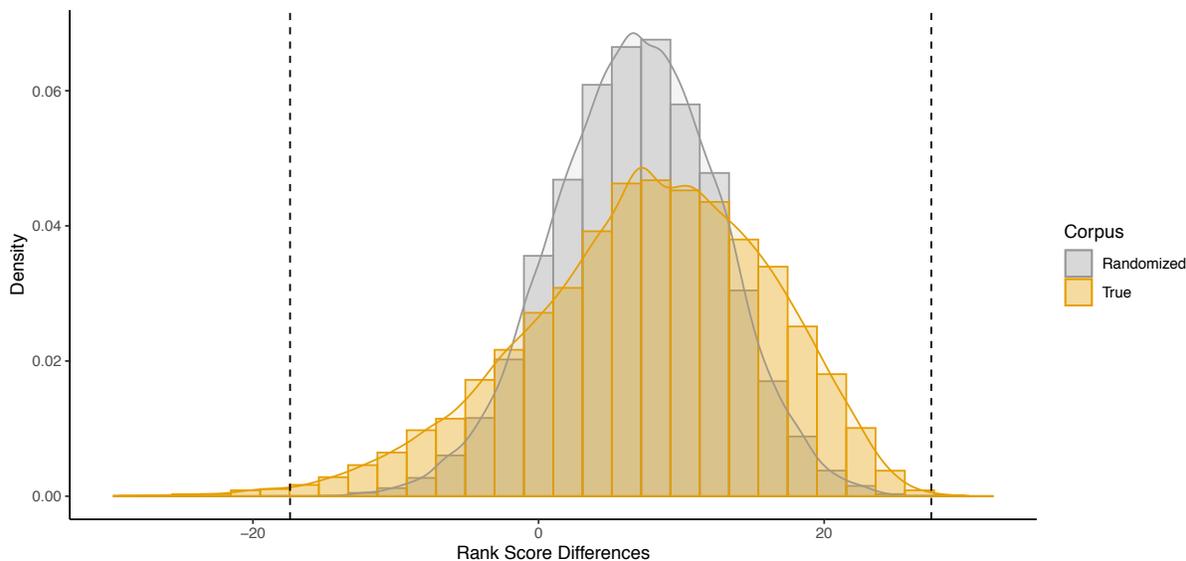
1. Compute share of counts of an n-gram v by a news outlet j : $sh_{vj} = \frac{\text{count}_{vj}}{\sum_v \text{count}_{vj}} \forall v, j$;
2. For each v , rank sh_{vj} across the news outlets $j \in \{1, \dots, 48\}$:
 $\text{rank}_{vj'} = 1$ if $sh_{vj'} = \max_j(sh_{vj})$
 $\text{rank}_{vj''} = 2$ if $sh_{vj''} = \max_{j:j \neq j'}(sh_{vj})$
 etc.;
3. Compute an average rank for each v and outlet type: $\text{Rank}_v^x = \frac{\sum_{j \in x} \text{rank}_{vj}}{\sum_{j \in x} 1}$;
4. For each v , compute the difference in ranks between the GC and independent news outlets, $\Delta \text{Rank}_v^{\text{GC-Ind}} = \text{Rank}_v^{\text{GC}} - \text{Rank}_v^{\text{Ind}}$;

The procedure above gives us the usage rankings for each word. While words with the highest and lowest $\Delta \text{Rank}_v^{\text{GC-Ind}}$ are natural candidates for being government-sensitive, it

²⁷We have also tested grouping named entities in topics by running LDA (Blei et al., 2003) and doing the analysis on the resulting topics. Like the results of the PCA, LDA topics include incidental words such as journalist names.

is not clear if these differences in usage can occur by chance, as well as how many words with high and low $\Delta\text{Rank}_v^{\text{GC-Ind}}$ we should classify as sensitive. To define the thresholds of unusually high and low rank score differences, we repeat steps 1-4 K times with randomly permuted word counts within an outlet. Each iteration gives us a random draw of the lowest $\Delta\text{RandomRank}_v^{\text{GC-Ind}}$. We consider a word to be significantly underused by the GC outlets if its $\Delta\text{Rank}_v^{\text{GC-Ind}}$ is below the average $\min_v \Delta_k^{\text{RandomRank}} \text{Rank}_v^{\text{GC-Ind}}$ across 1,000 permutations.²⁸ Online Appendix E.1 provides a more detailed exposition of the procedure and examples.

Figure 3: Histogram of differences in named entities usage between the GC and independent outlets.



Histograms are based on $\Delta\text{Rank}_v^{\text{GC-Ind}}$ and $\Delta\text{RandomRank}_v^{\text{GC-Ind}}$ rank score differences. Gold color corresponds to the actual corpus, silver color – to a random corpus. Vertical lines are the cutoff values for significantly under- or overused words, computed using $K = 1,000$ iterations.

Figure 3 presents the histogram of rank score differences $\Delta\text{Rank}_v^{\text{GC-Ind}}$ for all 38,790 common unigrams and bigrams of named entities. The distribution based on the actual corpus is in gold color; silver color corresponds to one random corpus draw. The actual corpus distribution has higher variance and a longer left tail, implying that there is a set of named entities that are systematically omitted by the GC outlets. We find 208 unigrams and bigrams of named entities that are significantly underused by the GC outlets; in contrast, only 14 named entities are overused by the GC outlets, and most of them are sports-related.²⁹

²⁸We run robustness tests with more and less restrictive thresholds. Our conclusions are unchanged.

²⁹Ten out of fourteen are names of hockey players and hockey teams (e.g. “Boston Bruins”, “Pittsburgh

Table 5: List of the top 20 unigrams and bigrams of named entities underused by GC news outlets.

Underused named entity: English translation	Information about the named entity	Rank Difference $\Delta\text{Rank}_v^{\text{Ind-Gov}}$
Rotenberg	Businessman, close ally of Vladimir Putin	-28.9
Roskomnadzor	Federal Agency exercising media censorship	-28.2
Khodorkovsky	Opposition, political prisoner	-28.1
Alexey Navalny	Opposition politician	-26.9
Navalny	Opposition politician	-26.5
Lebedev	Associate of Khodorkovsky, political prisoner	-25.5
Sechin	Head of Rosneft, close ally of Vladimir Putin	-25.5
Kudrin	Head of the Committee of Civil Initiatives	-25.3
Kosenko	Arrested at the opposition rally at Bolotnaya	-24.9
Sergei Guriev	Economist, interrogated about “Yukos”	-24.9
Bolotnaya	Place of a large opposition rally	-24.8
Prokhorov	Businessman, political activist at the time	-24.8
Bukovsky	Political activist	-24.7
Marat Gelman	Gallerist, fired for a political exposition	-24.7
Gennady Timchenko	Businessman, close ally of Vladimir Putin	-24.3
Sakharova	Place of a large opposition rally	-24.3
Svetlana Davydova	Civilian, investigated for treason	-24.3
Ketchum	PR agency working for Russian government	-24
Mikhail Khodorkovsky	Opposition, considered a political prisoner	-24
Gelman	Gallerist, fired for a political exposition	-23.9

We screen out named entities that relate to the profession of journalism – which may show up in the list simply due to news source citations – and get the remaining 128 named entities systematically underused by the GC outlets.³⁰ Table 5 presents a list of top 20 named entities that are the most underused by the GC outlets. All of these named entities are related to issues sensitive for the Russian government, such as political opposition (for instance, “Khodorkovsky” and “Navalny”), political protests (“Bolotnaya” and “Sakharova”), alleged corruption (“Rotenberg” and “Gennady Timchenko”) and media control (“Roskomnadzor” and “Ketchum”). The same pattern holds for the rest of the top 128 underused named entities, listed in Tables A1–A5 in the Online Appendix E. We classify any article that

Penguins”); the other four include a Russian astronaut (“Gennady Padalka”), a Russian missile system (“Pantsir”), a pro-Russia Polish activist (“Mateusz Piskorski”), and a bi-gram “Kiev Donbass”, a particular way to refer to the Ukraine crisis.

³⁰We use three independent research assistants to find named entities related to journalism, as well as ambiguous named entities. Online Appendix E.2 provides more details on the procedure.

mentions one of the top 128 underused named entities as covering government-sensitive news, and label those as “political protests, opposition and corruption” news, “POC” news for short.³¹

We validate our detection of sensitive news topics by a classification of named entities by three independent research assistants. The research assistants were tasked to rate named entities on a five point scale, ranging from “never related to sensitive news” (score of 1) to “always related to sensitive news” (score of 5). Based on the ratings of 200 named entities that are detected as the most underused by the GC outlets, our classification substantially outperforms all alternative methods, such as the comparison of named entity usage shares, TF-IDFs, partial least squares used by Gentzkow and Shapiro (2010), article-level Lasso regression (Tibshirani, 1996) and article-level naive Bayes classification. An average named entity classified as censored by our method got an average total score of 8.86 (out of 3 research assistants * 5 = 15 possible), compared to 5.54–7.13 average scores for the average censored named entity from the rest of the methods. Online Appendix E.2 describes the validation procedure and its results in detail, including Figure A2 that presents the average sensitivity scores across the six methods.

We further check whether there is difference in framing of the POC news on the GC and independent news outlets. For this, we compare usage of words that are not named entities in the articles classified to cover POC news topics. We use the same classification algorithm as described above. Figure A4 in Online Appendix E.4 presents the distribution of rank score differences. We find little evidence of framing in the POC news coverage – out of 34,688 words that are not named entities in these articles, 36 and 22 are classified as under- and overused by the GC outlets, respectively, and these numbers drop to 14 and 5 if we use a more strict threshold for detecting the underused words.³² The underused words include acronyms – misclassified as not named entities – that refer to the entities related to the censored topics.³³ Out of the few remaining words, the GC outlets underuse words related to court hearings and political arrests, such as “otkaz” (denial), “uznik” (prisoner), “specpriemnik” (detention center), “arest” (arrest), and “dopros” (interrogation),³⁴ and overuse words

³¹We additionally validate the 128 word threshold by examining the sensitivity score of each named entity assigned by three independent research assistants. Figure A3 in the Online Appendix E.2 presents the results.

³²A more restrictive threshold keeps only words with $\Delta\text{Rank}_v^{\text{GC-Ind}}$ below a 5% quantile of $\min_v \Delta_k^{\text{Random}} \text{Rank}_v^{\text{GC-Ind}}$ across 1,000 simulations.

³³For instance, “RBC”, a news outlet name, and “ECHR”, The European Court of Human Rights that often reviews cases against opposition in Russia.

³⁴Words are transliterated from Cyrillic.

related to sports, such as “sportsman”, “snowboarder”, and “champion”.³⁵ We conclude that, with an exception of the broader coverage of the process of political arrests, there is limited evidence of framing in the POC news, meaning that censorship is the main strategy of the controlled media.

4.1.2 Ideologically Framed News: The Ukraine Crisis

Apart from censored sensitive news about political protests, opposition and corruption that we have detected above, the main sensitive news topic in Russia in 2013-2015 was the Ukraine crisis of 2013-2014, with a subsequent conflict between Russia and Ukraine. The conflict was widely covered in the Russian news media with the reporting allegedly heavily slanted by the GC news outlets (themoscowtimes.com, 2014; time.com, 2014). The Ukraine crisis is also the only major topic (outside of the POC news) that is classified as sensitive by three independent research assistants – out of a randomly selected 724 named entities that were classified, 29 were given a high sensitivity score by the research assistants, with 79% of them (23 out of 29) being related to the Ukraine conflict.³⁶

Figure A5 in Online Appendix E.5 presents the share of news articles that contain the word “Ukraine” in the coverage of the GC, independent and potentially-influenced news outlets. Before the Ukraine crisis, an average of 2-3% of news articles mentioned Ukraine across all the news outlets. After February 22, 2014, the day Ukrainian president Yanukovich fled to Russia and the crisis unfolded, an average of 20-30% of news articles mention Ukraine, with the GC outlets systematically covering 5 percentage points more news about Ukraine compared to the independent outlets. This disproportional coverage suggests that censorship was not the primary strategy of the GC news outlets in handling the Ukraine crisis; if anything, they report more news about the Ukraine compared to other outlets. We classify any news article that mentions Ukraine as news about the Ukraine crisis.³⁷

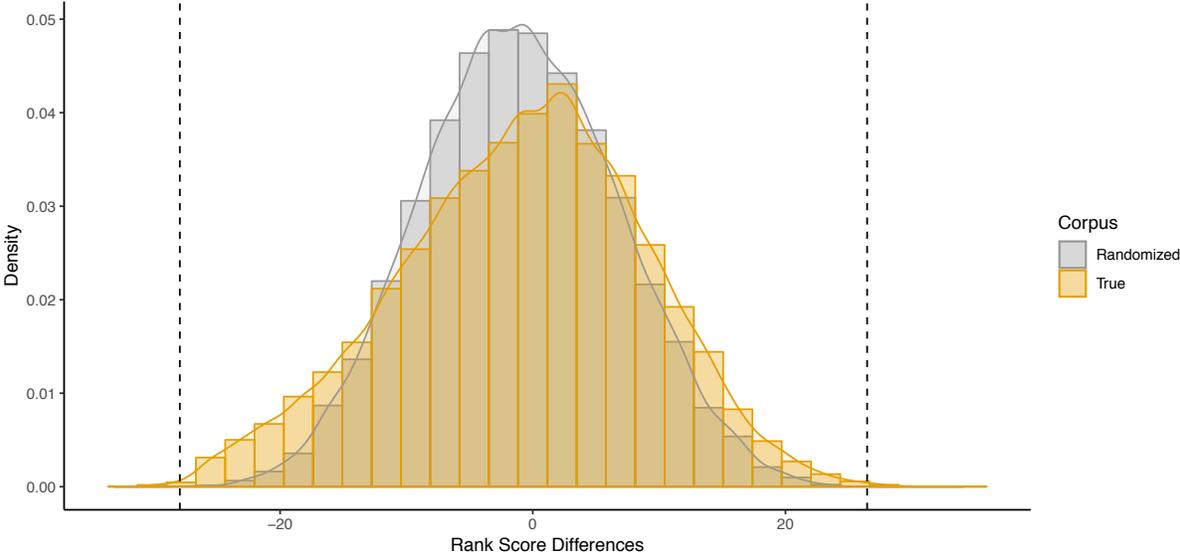
We check for framing in the Ukraine crisis news by comparing usage of words that are

³⁵All under- and overused words are presented in Table A7 in the Online Appendix E.4.

³⁶The five named entities that got the highest sensitivity score are “News Donbass”, “Euromaidan”, “Maidan”, “Donbass” and “Kiev Donbass”, all related to the Ukraine revolution that happened on Maidan Nezalezhnosti (Independence Square) in Kiev and that was followed by a war in Donbass, an area in the Eastern Ukraine. Online Appendix E.3 presents the details, with Table A6 presenting all 29 sensitive named entities.

³⁷We use this classification to keep the definition broad and ensure that we do not miss any articles related to the conflict. Alternatively, we can define news articles as being about the Ukraine crisis using 23 sensitive named entities detected by the research assistants, which we list in Table A6. Our results are robust to using this alternative classification. The correlation in the volume of the Ukraine-crisis news based on these two measures is 91.5%.

Figure 4: Histogram of differences in usage of non-named entities in the Ukraine crisis news, between the GC and Ukrainian outlets.



Histograms are based on $\Delta\text{Rank}_v^{\text{GC-Ukr}}$ and $\Delta\text{RandomRank}_v^{\text{GC-Ukr}}$ rank score differences. Gold color corresponds to the actual corpus, silver color – to a randomized corpus. Vertical lines are the 5% and 95% cutoff values for significantly under- or overused words, computed using $K = 500$ iterations.

not named entities by the GC and Ukrainian news outlets. Figure 4 presents the distribution of rank score differences. The shapes of the distributions of $\Delta\text{Rank}_v^{\text{GC-Ukr}}$ and $\Delta\text{RandomRank}_v^{\text{GC-Ukr}}$ are drastically different, showing systematic differences in the word usage. Out of 34,395 words in the corpus, we find 27 words that are significantly underused by the GC news outlets compared to the Ukrainian outlets, and 101 words that are significantly overused by the GC news outlets. The language underused by the GC news outlets includes an “annexation” (rank 3) and “occupation” (rank 10) of Crimea by Russia via a “pseudo-referendum” (rank 4), and a description of the Ukraine military that conducts an “anti-terroristic” (rank 5) operation against “separatists” (rank 13) in the Eastern Ukraine. In contrast, the GC news outlets describe the same events as a “reunion” (rank 1) of Russia with Crimea, and state that the Ukraine military conducts a “punitive” (rank 3), “russophobic” (rank 10), and “anti-Russian” (rank 18) operation in the Eastern Ukraine. We hired three independent research assistants to screen out incidental words that occur due to a broader difference in issues covered by Russian and Ukrainian news outlets. A final set of words that are labeled as having pro-Russia or pro-Ukraine slant by at least two research as-

sistants include 7 words underused and 26 words overused by the GC outlets.³⁸ We label any article that mentions one of these words as having a pro-Russia or pro-Ukraine ideological framing.

We further validate the detected ideological framing in three ways. First, the language detected by our procedure is remarkably consistent with the pro-Russian and pro-Ukraine propaganda narrative described by journalists and fact-checking websites (stopfake.org, 2014) – the pro-Russian slant frames Ukraine as a “fascist junta” that conducts a “punitive operation” in the Eastern Ukraine, and the pro-Ukraine slant frames Russia as an “aggressor” that has “occupied” the territory of the Ukraine.

Second, we get a similar measure of the ideological slant if we use a more restricted definition of the Ukraine crisis news articles – this removes most of the incidental words while keeping all of the words labels as sensitive by the research assistants. The correlation in the implied ideological positions of the news outlets is 0.968.

Finally, we run an additional validation by asking two independent research assistants to code up 1,075 news articles about the Ukraine crisis as having a pro-Russia or pro-Ukraine slant. The implied ideological positions of the news outlets computed with our measure match the ideological positions based on the classification by the research assistants, with the correlation of 0.839.³⁹

4.2 Coverage of Government-Sensitive News

Above, we have identified two government-sensitive news topics in the online news market in Russia – the POC and Ukraine-crisis news. We now leverage the knowledge of these topics and the corresponding framing to construct a measure of relative importance of news on a given day, as well as characterize the reporting of the news outlets.

The core idea behind our empirical strategy is that different days have different relative importance of sensitive news, depending on which sensitive events have happened that day, V_t^S . We recover the relative importance of news topics by computing the share of news articles covering this topic on a given day, $V_t^l = \frac{\sum_j N_{tj}^l}{\sum_l \sum_j N_{tj}^l}$, where N_{tj}^l is the number of articles outlet j writes about topic l on day t . On an average day, 9.56% of news articles cover the POC news, and 19.13% – news about the Ukraine crisis. There are large differences in the share of coverage across days, with the standard deviation of V_t^l of 3.75 and 11.3 percentage

³⁸We report the full list of under- and over-used words, as well as results of word classification, in Tables A8 and A9 in the Online Appendix E.6.

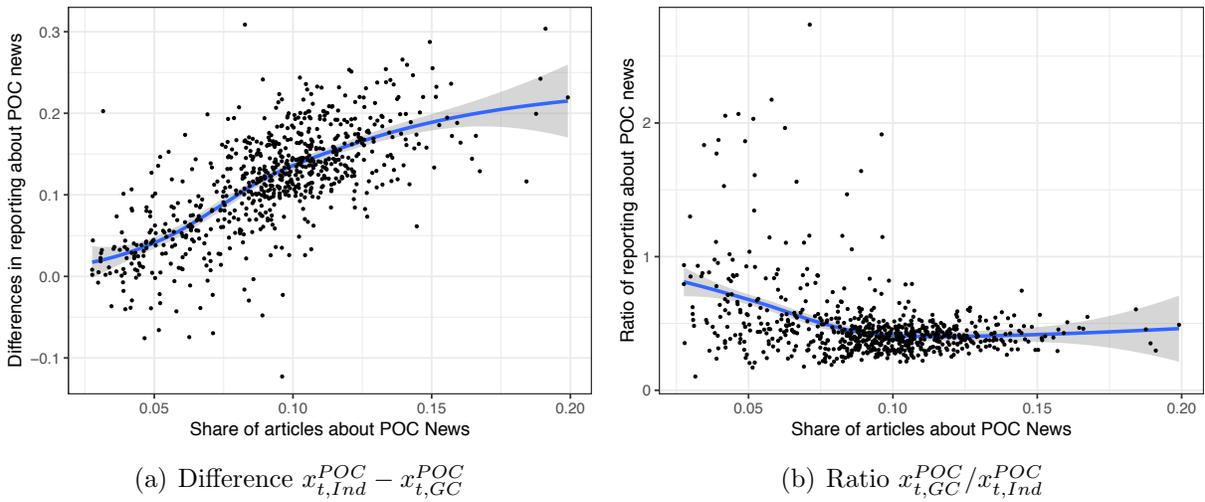
³⁹Figure 7 in Section 4.2.2 presents the ideological positions of the news outlets, and the Online Appendix E.7 further details the classification and validation procedure.

points, respectively. The implied coefficients of variation are 0.39 for the POC news and 0.59 for the Ukraine crisis news.⁴⁰

4.2.1 Coverage of POC News

News outlets decide how much news of each type to cover on day t . We measure this decision by computing the share of news articles covered by an outlet j on topic l on day t , $x_{tj}^l = \frac{N_{tj}^l}{\sum_l N_{tj}^l}$. In particular, for the censored POC news, we are interested in whether the difference in coverage of the GC and independent news outlets is higher on days with a higher relative importance of POC news, as predicted by the model in Section 2.1. Subfigure (a) in Figure 5 tracks the relationship between the difference in coverage of POC news by the GC and independent outlets, $x_{t,Ind}^{POC} - x_{t,GC}^{POC}$, and the relative importance of POC news, V_t^{POC} . There is a strong positive correlation between the two, confirming that censorship becomes more binding on days with more sensitive news events.

Figure 5: Differences in the POC news reporting by the GC and independent news outlets.



Subfigure (a) plots the relationship of $x_{t,Ind}^{POC} - x_{t,GC}^{POC}$ and V_t^{POC} , and Subfigure (b) plots the relationship of $x_{t,GC}^{POC}/x_{t,Ind}^{POC}$ and V_t^{POC} . The blue line corresponds to the fitted local polynomial regression.

Subfigure (b) in Figure 5 tracks shares of POC news that the GC outlets report compared

⁴⁰An alternative way of characterizing the relative importance of a news topic on a given day would be to collect information on the number of topic-related events that happened on this day. Using such an approach requires access to the full event data – which we do not have – and judging on whether a particular event is newsworthy or not. Using the relative importance solves the latter problem since we rely on news outlets’ decisions on whether an event is worth covering or not.

to the independent outlets, $\frac{x_{t,GC}^{POC}}{x_{t,Ind}^{POC}}$, in relation to the relative importance of POC news, V_t^{POC} . Except for the days with very few POC news events – where the results are noisy – the ratio $\frac{x_{t,GC}^{POC}}{x_{t,Ind}^{POC}}$ is a stable 0.42, meaning that on most days GC outlets cover 42% of POC news that the independent outlets cover. Importantly, this ratio does not change with shifts in the V_t^{POC} .

We further examine whether the share of the POC news reporting of each individual outlet changes with the relative importance of sensitive news V_t^{POC} . We find a very low correlation in the reporting shares x_{tj}^{POC}/V_t^{POC} and V_t^{POC} ; outlet fixed effects explain 30.05% of the variation in x_{tj}^{POC}/V_t^{POC} , while adding an interaction of outlet fixed effects and V_t^{POC} increases the R-squared only to 32.8%. Given such limited reaction of the news outlets to V_t^{POC} , we conclude that we can approximate the ideological positions of news outlets by their average share of reporting about the POC news, $\bar{x}_j^{POC} = \frac{\sum_t N_{tj}^{POC}}{\sum_t \sum_i N_{tj}^{POC}}$. Figure A6 in Online Appendix F presents the resulting ideological positions of the news outlets.

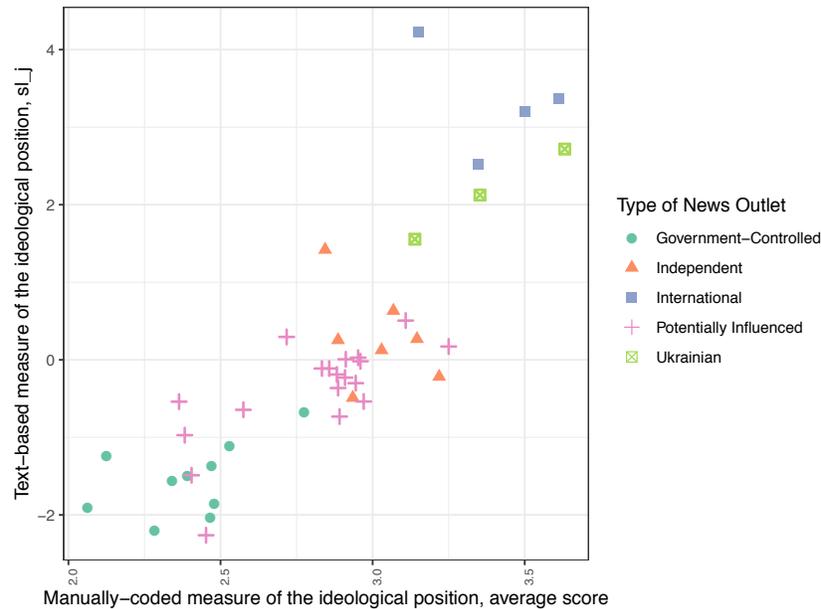
4.2.2 Coverage of Ukraine Crisis News

The primary method of government control in the Ukraine crisis news is the ideological framing of the news. We measure the ideological positions of the news outlets by taking the difference in shares of articles with a pro-Russia and pro-Ukraine slant that we have detected in Section 4.1.2. On day t for the news outlet j , the ideological framing of the Ukraine crisis is measured as $\frac{N_{tj}^{\text{pro-Russia}}}{N_{tj}^{\text{Ukr}}} - \frac{N_{tj}^{\text{pro-Ukraine}}}{N_{tj}^{\text{Ukr}}}$, where $N_{tj}^{\text{pro-Russia}}$ and $N_{tj}^{\text{pro-Ukraine}}$ is the number of articles with the pro-Russia and pro-Ukraine slant, and N_{tj}^{Ukr} is the number of articles about the Ukraine crisis. These ideological positions do not change with the volume of news about the Ukraine crisis, V_t^{Ukr} ; using the data since the beginning of the Ukraine crisis, outlet fixed effects explain 40.61% of the variation in the ideological positions, while adding an interaction of outlet fixed effects and V_t^{Ukr} explain only an additional 0.47 percentage points of the variation. Similarly, there is a limited difference in the share of the Ukraine crisis news that news outlets report; outlet fixed effects explain 61.69% of the variation in the share of reporting, and interactions of outlet fixed effects and V_t^{Ukr} increase R-squared to 67.9%.

Given the stability of the ideological positions, we approximate them with the difference in the overall share of articles with the pro-Russia and pro-Ukraine slant for each outlet. First, we compute $sl_j^{\text{pro-Russia}} = \frac{\sum_t N_{tj}^{\text{pro-Russia}}}{\sum_t N_{tj}^{\text{Ukr}}}$ and $sl_j^{\text{pro-Ukraine}} = \frac{\sum_t N_{tj}^{\text{pro-Ukraine}}}{\sum_t N_{tj}^{\text{Ukr}}}$, the average shares of using pro-Russia and pro-Ukraine slant for each outlet. On average, 7.2% of news articles about the Ukraine crisis use the language with a pro-Ukraine framing, and 21.8% use the language with a pro-Russia framing – partially because we detect more pro-Russia

expected, Ukrainian news outlets cover more news about the Ukraine crisis, followed by the international and GC news outlets.⁴¹

Figure 7: Text-based versus manually-coded ideological positions of news outlets in their Ukraine crisis coverage.



Each dot represents a position of a news outlet, with shapes and colors of the dots corresponding to the outlets' types. The correlation between the two measures is 0.839.

We validate the ideological positions of the news outlets by comparing them to the ideological positions implied by the manual classification of the news articles by two independent research assistants. Research assistants rated 25 articles about the Ukraine crisis for each news outlet in the sample, giving each article a score from 1 (heavy pro-Russia slant) to 5 (heavy pro-Ukraine slant). Figure 7 presents the resulting ideological positions of the news outlets based on the automatic sl_j measure and on the manually coded average scores. The correlation between the two measures of the ideological positions of the news outlets is 0.839, meaning that our text-based measure closely tracks the manually-labeled measure of ideological framing. Online Appendix E.7 presents further details of the classification and validation procedure.

⁴¹Figures A7 and A8 in Online Appendix F present the ideological positions and reporting of the news outlets with the corresponding outlet labels; Figure A9 presents a joint distribution of the share of POC news reported (censorship) and ideological framing of the Ukraine-crisis news (propaganda) across the news outlets.

4.3 Discussion of Sensitive News Detection

We pause for an additional discussion of our method of sensitive news classification.

First, the goal of our classification method is to separate out government-sensitive news topics, ones that all GC outlets publish in a systematically different way, from the rest of the news topics. This is different from a news outlet classification task in which we use text corpus to predict whether a news outlet is controlled by the government or not. In the latter task, we would need a method that detects deviations in language usage even if done by one or two GC outlets, since this information helps to predict their affiliation. In contrast, for our task we care about deviations in language usage that apply to all controlled news outlets. This is driven by our assumption that censorship strategy of the government applies to all the GC outlets. Such difference in objectives explains why our simple method outperforms the existing state-of-the-art classification methods described in Section 4 and Online Appendix E.2 – all of them are designed to predict the GC news outlets.

Second, while our measures of POC and Ukraine-crisis news are based on a small subset of named entities mentioned in the news, they proxy larger sensitive news topics. For instance, the named entity “Navalny” – a political activist and a prominent investigator and critic of corruption in the government – might come up in any news related to opposition and corruption, not only in the news stories about the investigations of Navalny. As a result, our measures of the volume of sensitive news and ideological positions of news outlets are robust to local changes in the number of words that describe sensitive news; for instance, if we manipulate the definition of POC news by moving around the cutoff from a more (89 censored named entities) to a less (400 censored named entities) restrictive measure, the implied measures of the POC news volume and reporting are almost unchanged – the average correlation in different measures of V_t^{POC} and \bar{x}_j^{POC} is 91% and 97%, respectively.

Third, we stop short of applying and validating our method against some more sophisticated text methods, both due to the infeasible computational intensity and due to our goal of separating out sensitive words from the incidental ones. Our sensitive news detection method can be applied to the news events or topics instead of individual words; for instance, one can detect news topics from word co-occurrence in the news articles (Blei et al., 2003) and then run our classification algorithm on an outlet-news topic matrix. The downside of this method is that topic detection methods will group informative and incidental words together, increasing the noise in the measure of government-sensitive news topics. Similarly, our method can be applied to word embeddings (Mikolov et al., 2013); in particular, structured exponential family embeddings (SEFE) developed by Rudolph et al. (2017) are

a natural fit to learn the difference in framing of particular named entities across the news outlets. Applying SEFE to a corpus of our size requires significant computational resources; we leave such extensions to future work.

Finally, we note that our classification of sensitive news is based on a comparison of topics published in the news market. In particular, we detect the degree of censorship by comparing news published by the GC and independent news outlets, which does not account for a potential self-censorship by the independent outlets. Schimpfossel and Yablokov (2014) discusses the reasons for self-censorship in the TV news market in Russia, and similar logic can be applied to the online news market. Our measure of censorship is thus closest to “state censorship” in the classification of Crabtree et al. (2015) applied to the Russian market.

5 Model-Free Evidence

Before estimating the empirical version of the model defined in Section 2, we present some model-free evidence that suggests the direction of consumer preferences and the source of demand for the GC news outlets.

5.1 Descriptive Evidence from News Consumption

We start with describing a typical news consumption process in the Russian online news market, and highlight the differences in news consumption on the GC and independent outlets – suggesting the potential drivers of the GC outlets’ consumption.

First, the data suggest that GC outlets benefit from third party referrals more than independent outlets. Table 6 splits the shares of referral traffic by outlet type, focusing on the first visit to a news outlet within a day. Direct navigation is the main source of traffic for all types of news outlets but plays a lower role for the GC outlets (50.68%), especially compared to the independent ones (56.28%). In contrast, the GC outlets get more than a quarter of their traffic from Yandex (25.57%), compared to only 15.5% for the independent outlets. The GC outlets also get a higher share of their traffic from other browsers, news aggregators, and news outlets in our sample, compared to the independent outlets.

The GC outlets further benefit from cross-referencing each other more than other types of the news outlets. Table 7 zooms in on the cross-referrals by news outlets in our sample, grouped by types. While the GC outlets are responsible for 34.3% of the outlet-to-outlet referrals, this share goes up to 68.09% if the landing website is of another GC outlet. In contrast, independent outlets get 68.93% of cross-referral traffic from the potentially influenced

Table 6: Summary of referral websites for first visit per day of news outlets, by outlet types.

Referral From	Outlet Type				
	GC	Potentially Influenced	Independent	International	Ukrainian
	Traffic Share (%)				
Direct & from This News Outlet	50.68	54.44	56.28	51.74	57.54
Yandex	25.57	22.78	15.50	15.66	14.64
Other Browsers (not Yandex)	6.76	8.62	5.63	16.17	9.78
Other Aggregators (not Yandex)	1.31	3.54	1.19	0.18	0.61
Other News Outlets	1.28	1.41	1.08	2.75	0.84
Social Media	0.40	0.33	0.24	0.39	0.58
Other Websites	14.00	8.90	20.07	13.10	16.00

The shares of traffic are computed conditional on the outlet type. Results are for the first visit of a news outlet on a given day. All columns sum up to 100%.

outlets, 20.71% from other independent outlets, and only 7.73% – from the GC outlets. Such differences in referrals are not driven by the switching patterns of consumers; the second part of Table 7 presents shares of first within-day outlet visits by types of outlets that were consumed right before on the same day. Looking across the columns, the shares of transitions in outlet-to-outlet switching are much more similar compared to the outlet-to-outlet referrals. In particular, consumers are much less likely to switch from one GC outlet to another (33.31%) than to be referred (68.09%), and much more likely to switch to the independent outlet from the GC outlet (30.39%) than to be referred (7.73%).

Second, the GC outlets get disproportionally more traffic from landings on the news articles and “other pages” than the rest of the outlets. Table 8 splits the shares of first website visits landing on different web pages by types of outlets. Pooling across all news outlets, 55.94% of first website visits land on the news article pages, and around 19.41% land on the main directory. For the GC outlets, the share of landings on news article pages is 58.28%, and only 10.94% land on the main pages. In contrast, 28.29% of the first visits of the independent outlets land on the main page. Further, landings on “other pages” account for 27.28% of all first daily website visits of the GC outlets, in contrast to 15.55% for independent outlets. This pattern is consistent with the importance of the third-party referrals for the GC outlets – consumers often skip the main page and are referred to the content of the GC outlets.

The high share of traffic navigating to “other pages” of the GC outlets reflects the availability of video content on these websites. In particular, 3 out of 10 GC news outlets in our sample are major federal TV channels, with some of them streaming their content online.

Table 7: Outlet-to-outlet referrals and switching patterns, by outlet types.

Coming From:	Outlet Type					
	All	GC	Potentially Influenced	Independent	International	Ukrainian
	Traffic Share (%)					
	Referral patterns					
GC	34.29	68.09	24.85	7.73	1.47	2.95
Pot. Influenced	35.70	27.50	23.67	68.93	83.44	9.86
Independent	26.40	3.97	49.43	20.71	10.09	5.20
International	1.43	0.29	1.86	2.35	2.32	0.67
Ukrainian	2.19	0.15	0.19	0.28	2.69	81.32
	Switching patterns					
GC	30.67	33.31	30.06	30.39	19.31	20.58
Pot. Influenced	43.13	42.06	42.71	46.27	47.12	35.10
Independent	22.03	21.86	23.53	19.08	24.63	16.37
International	2.03	1.36	1.98	2.69	4.01	4.86
Ukrainian	2.14	1.41	1.71	1.58	4.93	23.09

The shares of traffic are computed conditional on the outlet type, only for traffic that is referred to a news outlet by other news outlets in our sample. Results are for the first visit of a news outlet on a given day. All columns within the referral and switching blocks sum up to 100%.

The top 2 “other pages” of the GC outlets visited by consumers are live streams of Channel One (www.1tv.ru) and Russia24 (www.vesti.ru), the two main federal TV channels in Russia. The other 3 out of 5 top “other pages” of the GC outlets are the reruns of the TV programs on the website of Channel One. While only a minority – around 10%, both for GC and independent outlets – of consumers who land on the websites through non-news article web pages go on to read news articles on the websites, the difference in the share of arrivals through “other pages” suggests that GC outlets get some benefit from video content on their website in driving news article readership.⁴²

Third, we examine which news articles and topics capture the highest attention share of consumers. For this, for each news outlet, we compute the consumption and publication shares of news articles about the POC and Ukraine-crisis news. The consumption share is defined as the share of visits of articles about the POC or Ukraine-crisis news among the visits of all news articles.

Figure A11 in Online Appendix I presents the resulting consumption and publication

⁴²Further, a lot of articles of the GC outlets have a video on top of the page, before the article’s text.

Table 8: First visit shares by types of web pages.

Referral From	Outlet Type					
	All	GC	Potentially Influenced	Independent	International	Ukrainian
	Traffic Share (%)					
Main Page	19.41	10.94	21.53	28.29	14.16	24.01
News Articles	55.94	58.28	57.01	50.54	51.63	60.25
News Subdirectories	7.80	3.49	12.29	5.62	15.03	9.12
Other	16.85	27.28	9.18	15.55	19.18	6.62

The shares of traffic are computed conditional on the outlet type. Results are for the first visit of a news outlet on a given day. All columns sum up to 100%.

shares of news articles about the POC and Ukraine-crisis news. For both POC and Ukraine-crisis news articles, the consumption share is higher than the publication for the average outlet. Interestingly, for the POC news articles, the difference in the consumption and publication shares is larger for outlets with more POC news coverage (p-value of 0.045), suggesting that readers of outlets with fewer POC news articles (i.e., the GC outlets) are less interested in the POC news.

We further check how the consumption share of Ukraine-crisis-related articles depends on the ideological framing of the conflict by the outlet. We find that outlets with a more pro-Ukraine ideological position have a higher share of consumption of the Ukraine-crisis news articles – one standard deviation shift of the outlet’s ideological position towards a pro-Ukraine framing is correlated with 3.7 percentage points (p-value = 0.019) higher Ukraine-crisis news readership, suggesting readers’ preference for the pro-Ukraine ideological framing of the news.

To get a better understanding of what kind of articles consumers read on each outlet, Tables A13 and A14 in Online Appendix I list titles of the top read news article on each news outlet in our sample. More than half of the most-read news articles across the outlets (26 out of 46 outlets) are about the Ukraine crisis, and around 11% are related to the POC news. For the GC outlets, 5 out of 10 outlets have the most-read articles related to the Ukraine crisis, and the rest are related to foreign policy and celebrity news. Similarly, 3 out of 10 most-read articles on the independent outlets are related to the Ukraine-crisis news – but, in contrast to the GC outlets, another 4 out of 10 are related to the POC news. All most-read news articles of the international and Ukrainian outlets are related either to the Ukraine crisis or POC news.

Finally, we examine the probabilities of consumers to continue reading other news after

arriving for a particular news article topic, exhibiting the behavior consistent with choice inertia and outlet-level switching costs. Table A15 in Online Appendix I presents the share of sessions where a consumer arrived on a news article of a particular topic and visited other news articles after that. Most of the times, consumers read only one news article upon arrival – consumers continue to navigate to other articles only in one out of four cases. Arriving on an article of a particular topic increases the probability that the consumer will continue reading articles of this topic; for instance, consumers who arrive on POC news articles have a 10.1% probability of reading another article of this topic, compared to 3.96% and 7.05% probabilities if they landed on non-sensitive or Ukraine-crisis news, respectively. The probability of reading another article about POC news is lower for GC outlets (8.39%) and higher for the independent outlets (11.82%). Further, only 2.92% of consumers who arrive on the GC outlets for non-sensitive news go on to read POC-related news, whereas this share is 5.43% for the independent outlets.

5.2 Changes in Market Shares with Sensitive News

We now examine how market shares of news outlets change in response to shifts in the volume of sensitive news in the market, V_t^{POC} and V_t^{Ukr} . This relationship is the cornerstone of our identification strategy; the ideological positions of news outlets – such as the share of sensitive news reported, \bar{x}_j , and ideological framing, sl_j – become more important for consumers on the days when there are more sensitive news events. This implies that, all else equal, consumers are more likely to navigate to news outlets with their preferred ideological position on days with a large volume of sensitive news, disproportionately increasing their market shares.

We construct the market shares of the news outlets using the news consumption records in the IE Toolbar data. We define news consumption of an outlet j on day t by consumer i as a visit to any page on the news outlet j .⁴³ We define the outside option as consumer i browsing on day t but not visiting any news outlets. The market share of the news outlet j on day t is then defined as the sum of all consumptions of j at t , divided by the sum of all outlets' consumption counts and outside option choices on t .

Before we get to the analysis, we plot the market shares by outlet type in Figure A12 in Online Appendix J.1. While the market shares of news outlet types are fairly stable across

⁴³Our results are robust to alternative definitions of news consumption, such as (a) a visit to at least one news article on outlet j , (b) a visit to any page but the main directory, (c) a visit to at least 5 pages on website j , (d) spending at least 2 and 3 minutes on website j .

the weeks, the biggest change – which is especially visible in Subfigure (b) where we do not account for the outside option choices – happens in the end of February 2014, when the Ukraine crisis starts. That week, the market shares of all but GC outlet types increased – by 14.6%-15.5% for the independent and potentially influenced outlets and by 42.1%-42.4% for the international and Ukrainian outlets. In contrast, the market share of the GC outlets dropped by 4.9%. While these changes in the market shares are based on just two-week aggregated data points, it gives us some idea on what to expect as we zoom into the data.

We now examine the relationship between the market shares and the volume of sensitive news, V_t^{POC} and V_t^{Ukr} , more formally by running a separate log-log regression of market shares on the volume of sensitive news for each outlet,

$$\log(\text{share}_{jt}) = b_{0j} + b_j^{POC} \log(V_t^{POC}) + b_j^{Ukr} \log(V_t^{Ukr}) + b_j^{Plac} \log(V_t^{Plac}) + Z'_{jt} d_j + \xi_{jt} \quad (7)$$

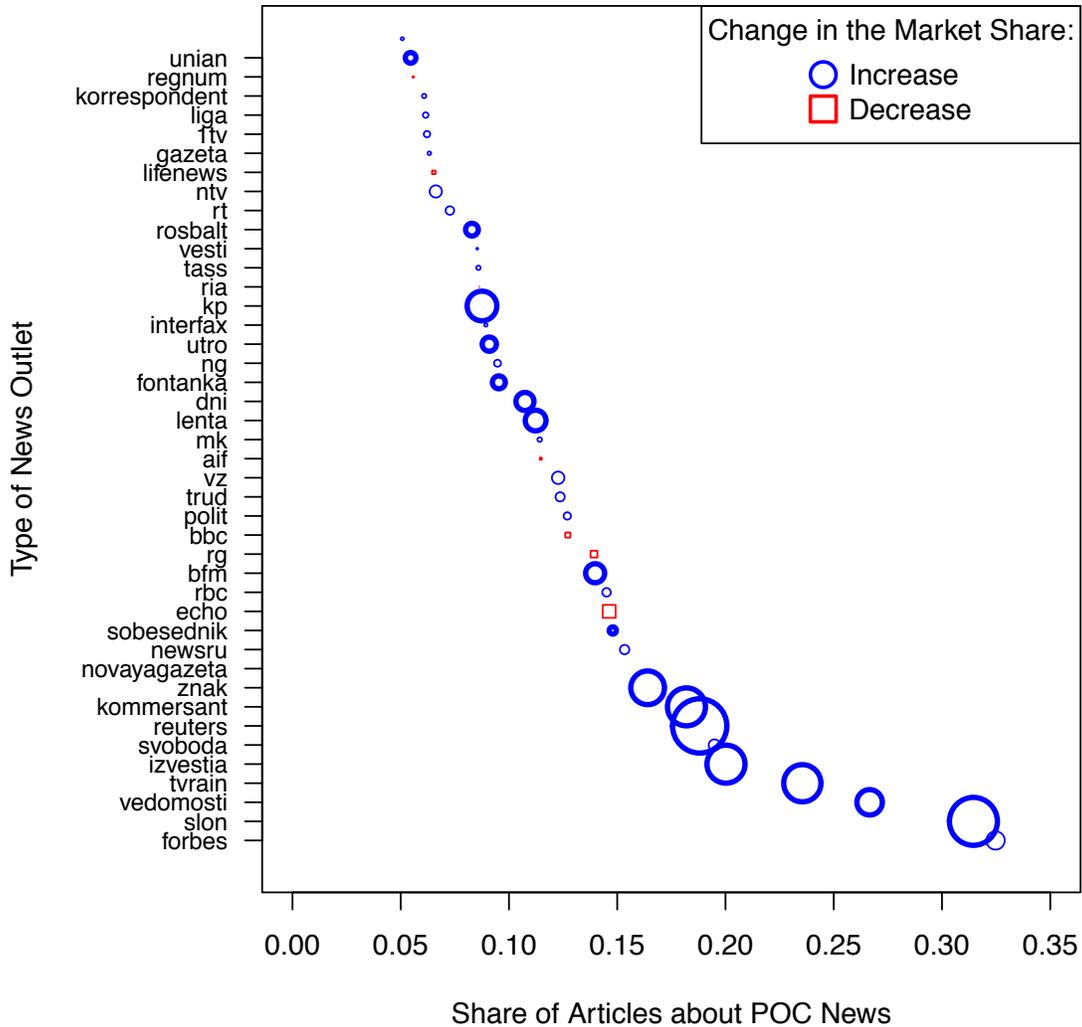
where Z_{jt} are controls – outlet-specific week and weekday fixed effects in the main specification.⁴⁴ The placebo variable, V_t^{Plac} , is the share of news articles on day t that mention one of 233 named entities that were coded by research assistants as not sensitive for the government (out of 724 randomly selected named entities). Since this measure includes random words related to different topics, we do not expect V_t^{Plac} to have any systematic effect on the market shares.

The slope coefficients, b_j^{POC} and b_j^{Ukr} , estimate the relationship between outlets' market shares and the volume of sensitive news in the market on day t . As long as the conditional independence assumption (CIA) holds, $\xi_{jt} \perp \log(V_t^l) | Z_{jt} \forall j, l = \{POC, Ukr\}$, we can interpret the estimates of b_j^{POC} and b_j^{Ukr} from regression 7 as causal effects of sensitive news volume on the outlets' market shares. CIA is a plausible assumption given that $\log(V_t^l)$ is determined by the number of sensitive news events that happen on day t , a process that is not controlled by the market participants.⁴⁵ Such reactions of market shares to V_t^{POC} and V_t^{Ukr} might be driven by consumers' preferences for the outlets' reporting and ideological positions.

⁴⁴Some outlets have no observed consumption on some days, leading to the market shares of zero. To avoid the problem of taking a logarithm of zero, we assign the lowest observed non-zero market share of this outlet to the days with zero consumption.

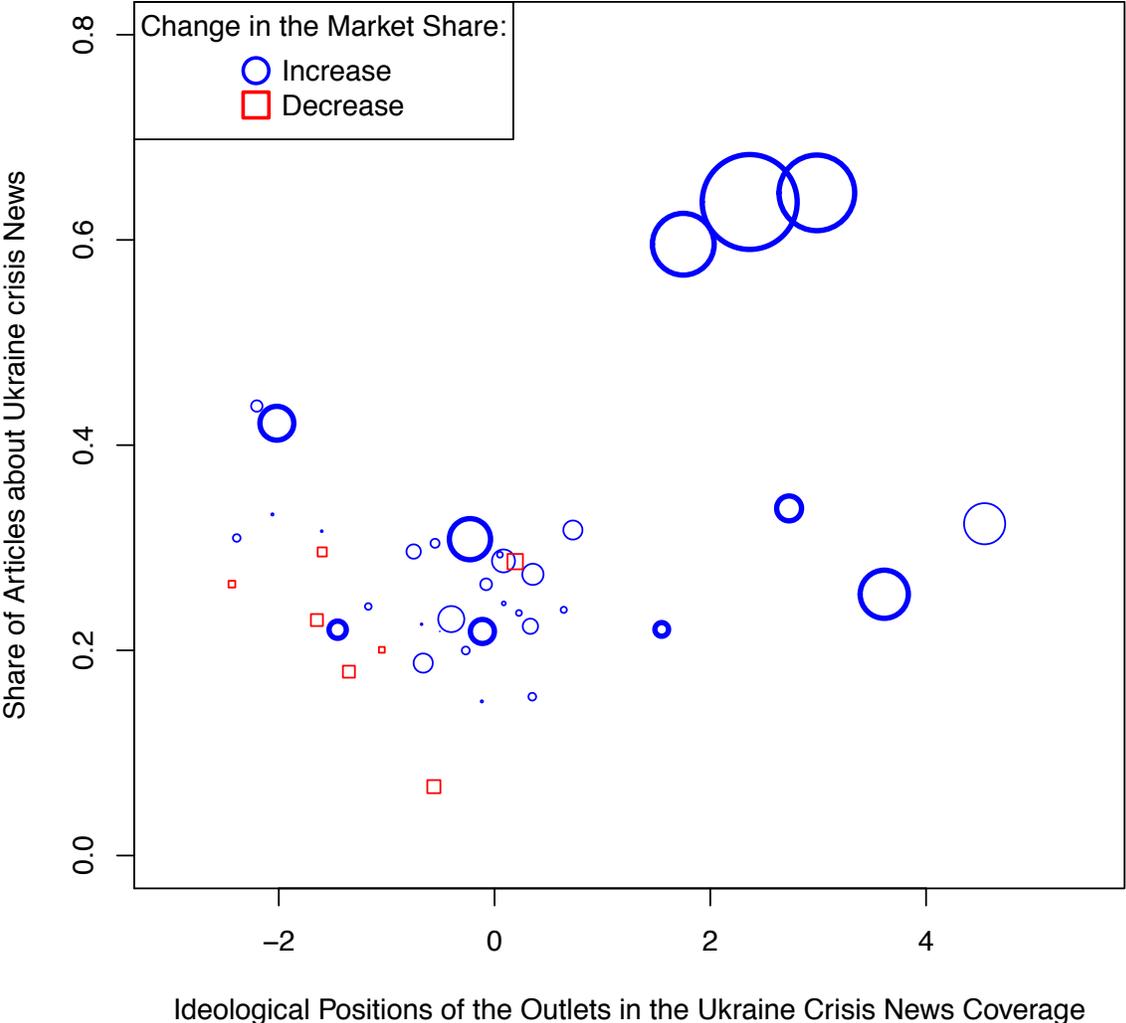
⁴⁵This assumption would be violated if the Russian government had control over all sensitive news events and was timing them strategically so that they overlap with some other significant news, similar to the strategic timing of the Israeli attacks on Palestine (Durante et al., 2015). We consider this unlikely, since in this context a lot of the sensitive news events are determined by other political actors (protests, corruption revelations, etc.). Moreover, even if the government has some control over the sensitive news events, the timing of these events is often influenced by other factors, such as the Ukrainian revolution, actions in the Eastern Ukraine, etc.

Figure 8: Estimates of correlations in the market shares of the outlets and relative importance of the Ukraine-crisis news, V_t^{POC} .



Each point represents a news outlet. The size of each point represents the effect of V_t^{POC} on the market shares of news outlets, measured in percentages (b_j^{POC} coefficient of regression 7). The blue color corresponds to positive coefficients, and the red color – to negative coefficients. The bold borders of the points correspond to significance of the change in the market share.

Figure 9: Estimates of correlations in the market shares of the outlets and relative importance of the Ukraine-crisis news, V_t^{Ukr} .



Each point represents a news outlet. The size of each point represents the effect of V_t^{Ukr} on the market shares of news outlets, measured in percentages (b_j^{Ukr} coefficient of regression 7). The blue color corresponds to positive coefficients, and the red color – to negative coefficients. The bold borders of the points correspond to significance of the change in the market share.

We estimate regression 7 for 42 news outlets in our sample.⁴⁶ Figures 8 and 9 visualize the estimates; each point represents an estimate of b_j^{POC} (Figure 8) and b_j^{Ukr} (Figure 9) for one of the 42 outlets. Points of larger size represent a larger absolute value of the estimates, with blue and red colors corresponding to positive and negative estimates. Points with bold borders represent outlets with the estimates significant at the 5% level.⁴⁷

News outlets in Figure 8 are ordered by the share of POC news they report. Seven out of nine news outlets with the highest share of reporting about sensitive news get a statistically significant increase in their market shares on days with a high $\log(V_t^{POC})$, and the other two are marginally significant at 10% level (p-values of 0.106 and 0.118). The average slope coefficient for these nine outlets is 0.219, meaning that a 1% increase in V_t^{POC} leads to a 0.22% increase in these outlets' market shares. In contrast, only 9 out of the other 33 outlets get a significant increase in their market shares on days with a high $\log(V_t^{POC})$, with an average b_j^{POC} estimate of 0.044. Figure A13 in Online Appendix J.2 plots b_j^{POC} estimates against news outlets' share of POC news reporting, \bar{x}_j^{POC} ; the relationship is positive and statistically significant.

Figure 9 presents estimates of b_j^{Ukr} , with the news outlets plotted by the share of Ukraine-crisis news they report, \bar{x}_j^{Ukr} , and their ideological framing, sl_j . News outlets that report a higher share of news about the Ukraine crisis, \bar{x}_j^{Ukr} , and have a more pro-Ukraine ideological position, sl_j , get the highest increases in their market shares on days with a high V_t^{Ukr} . In particular, six out of seven news outlets with the most pro-Ukraine slant get a statistically significant increase in their market shares, and the last one is marginally significant at 5% level (p-value of 0.056). The average slope coefficient for these seven outlets is 0.267, meaning that a 1% increase in V_t^{POC} leads to a 0.27% increase in these outlets' market shares. In contrast, only 4 out of the other 35 outlets get a significant increase in their market shares on days with a high $\log(V_t^{Ukr})$, with an average b_j^{Ukr} estimate of 0.041. Figures A14 and A15 in Online Appendix J.2 order the estimates by outlets' reporting and slant, \bar{x}_j^{Ukr} and sl_j , and Figures A16 and A17 plot b_j^{Ukr} estimates against news outlets' \bar{x}_j^{Ukr} and sl_j ; the relationship of b_j^{Ukr} with both \bar{x}_j^{Ukr} and sl_j is positive and statistically significant.

We do not find any systematic correlations of outlets' market shares with the placebo topic, $\log(V_t^{Plac})$; only 2 out of 42 outlets have significantly higher market shares on days with a high $\log(V_t^{POC})$, and another 2 have significantly lower market shares. Table A18

⁴⁶We exclude five news outlets for which we do not have information about the text of the articles, and one news outlet (*dw.de/ru*) for which we have few (10) news consumption occasions.

⁴⁷Tables A16–A17 and Figures A13–A17 in Online Appendix J.2 present estimates of b_j^{POC} and b_j^{Ukr} . Standard errors are heteroskedasticity and autocorrelation consistent.

and Figures A18 and A19 in Online Appendix J.3 present the estimates. We further check whether the estimates of b_j^{POC} and b_j^{Ukr} are correlated with the outlets' share of reporting about the placebo news topic, \bar{x}_j^{Plac} , and find both relationships to be statistically insignificant. Overall, all placebo tests confirm that the relationships we describe in Figures 8 and 9 are not incidental.

We confirm the relationship between the market shares and reporting and ideological positions of the news outlets by running a joint regression for all outlets in our sample,

$$\begin{aligned} \log(\text{share}_{jt}) = & b_{0j} + \log(V_t^{Ukr})(b^{Ukr} + d^{Ukr}\bar{x}_j^{Ukr} + d^{sl}sl_j) + \\ & + \log(V_t^{POC})(b^{POC} + d^{POC}\bar{x}_j^{POC}) + \log(V_t^{Plac})(b^{Plac} + d^{Plac}\bar{x}_j^{Plac}) + Z'_{jt}d_j + \xi_{jt}. \end{aligned} \quad (8)$$

The coefficients of interest are d^{POC} , d^{Ukr} and d^{sl} , interactions of changes in the volume of sensitive news and outlets' ideological positions. We also include the placebo topic, $\log(V_t^{Plac})$, and the corresponding share of reporting about this topic, \bar{x}_j^{Plac} .⁴⁸ Such a setup is similar in logic to shift-share instruments (Bartik, 1991), in which our identification argument relies on the quasi-random assignment of shocks, $\{V_t^{Ukr}, V_t^{POC}\}$, with potentially endogenous shares, $\{\bar{x}_j^{POC}, \bar{x}_j^{Ukr}, sl_j\}$ (Borusyak et al., 2018). Standard errors are clustered two-ways on the week and outlet level (Cameron et al., 2011).

Table 9 presents the regression results with different levels of fixed effects. Our preferred specification (3) includes outlet-specific week and weekday fixed effects. The joint regression estimates confirm our conclusions from the outlet-by-outlet market share regressions; market shares of outlets with higher \bar{x}_j^{POC} gain extra market shares on days with a high V_t^{POC} , and market shares of outlets with higher \bar{x}_j^{Ukr} and more pro-Ukraine sl_j gain extra market shares on days with a high V_t^{Ukr} . News outlets that report more about the placebo topic, \bar{x}_j^{Plac} , do not gain extra market share on days with higher $\log(V_t^{Plac})$.

We further check the robustness of our results to alternative consumption specifications. Results hold if we define news consumption by consumer i on day t as (a) a visit to at least one news article on outlet j , (b) a visit to any page of j but the main directory, (c) a visit to at least 5 pages on website j , and (d) spending at least 3 minutes (median time spent in the data) on website j . We also get similar results if we run the analysis separately for frequent and infrequent news consumers.⁴⁹

⁴⁸We exclude one outlet (“znak”) which is an outlier in terms of share of the placebo news topic covered, \bar{x}_j^{Plac} . See Figure A19 in in Online Appendix J.3. Table A19 in the same Online Appendix presents regression results with this website and without the placebo variables. All results hold.

⁴⁹We define frequent news consumers as people who read news on at least 10 days in our sample.

Table 9: Estimates of regression 8.

	$\log(\text{share}_{jt})$		
	(1)	(2)	(3)
b^{POC}	0.029 (0.046)	0.004 (0.041)	-0.030 (0.035)
d^{POC}	0.972*** (0.331)	1.179*** (0.307)	0.827*** (0.198)
b^{Ukr}	-0.184* (0.097)	-0.185*** (0.050)	-0.096** (0.049)
d^{Ukr}	1.179*** (0.310)	0.766*** (0.124)	0.600*** (0.136)
d^{sl}	-0.001 (0.027)	0.025** (0.011)	0.024** (0.010)
b^{Plac}	-0.105 (0.100)	-0.065 (0.056)	-0.011 (0.060)
d^{Plac}	1.587 (1.360)	0.859 (0.890)	0.402 (0.586)
Controls:			
Weekday FE	N	N	Y
Week FE	N	Y	Y
Observations	21,084	21,084	21,084
R ²	0.918	0.973	0.977
Adjusted R ²	0.918	0.969	0.973

*p<0.1; **p<0.05; ***p<0.01

Standard errors are clustered two-way on week and outlet level.

Overall, we have established that on days with a lot of sensitive news reporting, news outlets with more sensitive news reporting and more anti-government slant in the Ukraine-crisis get gain the most in their market shares. As long as the CIA holds, we can interpret these market share changes as causal effects of the volume of sensitive news on the market shares, which can be driven by consumer preferences for sensitive news reporting and anti-government slant. In this sense, the evidence points towards a preference of an average consumer for sensitive news topics and anti-government slant.

At the same time, the causal effects that we find might have alternative explanations. For instance, consumers' outlet and topic preferences might be correlated, and perhaps consumers with high persistent preferences for independent outlets – which cover a lot of sensitive news – are those who also prefer sensitive news topics over non-sensitive news. In this case, a disproportional market share increase of outlets that cover sensitive news might be driven by a small share of consumers who sort into the market on days with high V_t^{POC} and V_t^{Ukr} . Alternatively, consumers might navigate to anti-government websites to get a second opinion on the sensitive news, exhibiting conscientious consumption. To separate out such alternative explanations, we estimate a structural model of demand for news, accounting for potential consumer heterogeneity and conscientious consumption behavior.

6 Empirical Specification

In this section we bring together the stylized model from Section 2 and the empirical setting of the Russian online news market. We write down the empirical version of the model and describe the estimation procedure.

6.1 Empirical Model

There are I consumers and J news outlets in the market. On days when consumers spend time browsing online, they might choose to consume one or more news outlets, or decide not to read news from any of the outlets. Following Gentzkow and Shapiro (2015), we assume that consumers can read at most one news outlet at a time – it is impractical for people to read multiple news outlets simultaneously. This setting naturally lends itself to a discrete choice model, where on a consumption occasion τ a consumer chooses an outlet j that she has not read on the previous choice occasions $1, \dots, \tau - 1$ on this day. We define the 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 .⁵⁰ Thus, on each day t , consumers can have at most J news consumption occasions. Unless a consumer has read all J news outlets on day t , on the last τ of the day a consumer chooses an outside option of not consuming the remaining news outlets.

There are three news topics covered by the outlets: non-sensitive, POC and Ukraine-crisis news. The news event realizations are driven by a stochastic process that is not controlled by the market participants. The relative importance of each news topic over time is captured by overall share of news about this topic on this day, V_t^{POC} and V_t^{Ukr} , which we have defined in Section 4.2.

The J news outlets in the market make three decisions about the sensitive news reporting – which share of the POC and Ukraine-crisis news to report, and which ideological position to take in the reporting about the Ukraine-crisis news. The decisions are captured by the share of reporting of sensitive news, \bar{x}_j^{POC} and \bar{x}_j^{Ukr} , and the ideological framing in the Ukraine-crisis news, sl_j , which we define in Section 4.2. The importance of these ideological positions for consumer choice is shifted by the relative importance of the sensitive news on day t , V_t^{POC} and V_t^{Ukr} . Finally, outlets can also choose to differentiate in terms of their persistent features, such as which non-sensitive news to report and how much money to invest in quality of the news reporting or website, among others.

We take this empirical context to the model described in Section 2.2. At each choice occasion τ on day t , a consumer chooses an outlet j such that $u_{ijt\tau} \geq u_{ij't\tau} \forall j' \in \{0, \dots, J\} : j' \neq j$. We denote consumers' choices as $y_{it\tau}$. Adapting consumer utility defined in equation 6, we get

$$u_{ijt\tau} = \alpha_{ij} + V_t^{Ukr} \bar{x}_j^{Ukr} (\beta_i^{Ukr} + sl_j \gamma_i + |sl_j - s_{i\tau}|(\tau > 1)\rho_i) + V_t^{POC} \bar{x}_j^{POC} \beta_i^{POC} + |sl_j - s_{i\tau}|(\tau > 1)\eta_i + state_{it\tau}\pi_i + \epsilon_{ijt\tau}. \quad (9)$$

Equation 6 closely tracks the model defined in Section 2.2. Persistent preferences of consumers are defined by α_{ij} , a time-invariant taste of consumer i for outlet j . Consumers are allowed to hold a relative preference for POC and Ukraine-crisis news (over the non-sensitive news), captured by parameters β_i^{POC} and β_i^{Ukr} , respectively. Further, consumers have a preference for the ideological framing of the Ukraine-crisis news, captured by γ_i , and ρ_i is the variety-seeking parameter that shows whether consumers are more likely to read more ideologically-diverse news outlets on days with a lot of Ukraine-crisis events. The

⁵⁰This discrete-choice specification ignores the intensity of news consumption within the outlet; our results are robust to redefining a consumption occasion of an outlet to a visit to a news article on a given day, allowing for multiple articles read within a day.

reduced-form parameter η_i captures the baseline variety-seeking on days with no Ukraine-crisis news. The only new term compared to equation 6, $\text{state}_{it\tau}$, is an indicator variable that captures whether a consumer i has already visited j on day t . Since, by construction, consumers never revisit the same news outlet on day t , the variable $\text{state}_{it\tau}$ serves a technical purpose of restricting the actual choice set of consumers (with a highly negative value of π_i).

6.1.1 Discussion of the Assumptions

We pause to discuss several assumptions underlying this empirical model.

First, we assume that consumers know the relative importance of news topics on day t , V_t^{POC} and V_t^{Ukr} , and the reporting and ideological positions of the news outlets, \bar{x}_j^{POC} , \bar{x}_j^{Ukr} , and sl_j . We believe that these are reasonable assumptions in our context. We define consumption as visits to news articles, meaning that consumers have some exposure to the overall set of topics that have happened on day t , either on the main page of news outlets or on news aggregators. Our estimation also focuses only on frequent news consumers, who are more likely to know the average reporting positions. If these assumptions are violated, we likely overestimate the role of the persistent preferences of consumers and underestimate the preferences for the news reporting and ideological framing of sensitive news.

Second, we assume that consumer preferences for news topics and ideological framing of the news are stable over time. If this assumption is violated, our estimates would capture only average and short-term preferences of consumers. In particular, the estimates of persistent preferences, α_{ij} , capture any long-term effects of the ideology of the news outlets, as well as any unobserved differences in the sensitive news coverage other than the coverage of the POC and Ukraine-crisis news.

Third, we follow the stylized model and define the consumers' tastes for sensitive news topics as a coefficient on $V_t^{Sens}\bar{x}_j^{Sens}$. An alternative model specification is to separate out the effect of V_t^{Sens} , the relative importance of sensitive news on this day, and the effect of \bar{x}_j^{Sens} , the share of news on the outlets' website devoted to this topic. While separately identifying the effect of sensitive news coverage, \bar{x}_j^{Sens} , is appealing, such alternative specification makes it hard to identify and interpret consumers' tastes for sensitive news. In particular, in this alternative specification, the model needs to estimate not only coefficients on V_t^{Sens} and \bar{x}_j^{Sens} , but also a correlation term between them – pushing the requirements on the number of choices observed per consumer. Separately, such alternative specification deviates from the stylized model defined in Section 2, making it hard to interpret the coefficient estimates.

Finally, our model does not allow for the interactions between the volume of news coverage

of a topic and the quality of this topic. Any horizontal or vertical differences across the news outlets are captured by the persistent preferences of consumers, α_{ij} . In particular, if outlets have different quality of the non-sensitive news coverage, the differences are captured by α_{ij} .

6.2 Estimation

We use only frequent news consumers – those who consume news at least 10 days in our data sample period – to estimate the model. These consumers are more likely to be knowledgeable about the ideological positions of news outlets, and since they make more outlet choices, their data provides more information about potentially-heterogeneous preferences. There are 54,905 such news consumers in our sample.⁵¹ These news readers have 4,822,667 consumption occasions, or outlet-day visits. On almost half (48.6%) of the consumption days, news readers in the selected sample have only one news consumption occasion. However, conditional on having more than one consumption occasion on day t , news readers navigate to an average of 2.71 news outlets. For computational reasons, we estimate the model on a random sample of 10,000 of such frequent news consumers; all of our conclusions replicate if we re-run the model with a new random sample of consumers. As in Section 5.2, we focus on the top 42 online news outlets in the sample.

We estimate the distribution of $\theta_i = \{\alpha_{ij}, \beta_i^{Ukr}, \beta_i^{POC}, \gamma_i, \rho_i, \eta_i, \pi_i\}$ using a Bayesian hierarchical model. We assume that $\epsilon_{ijt\tau} \sim i.i.d. EV(0,1)$, leading to a standard logistic regression, but allow for a flexible heterogeneity in consumer preferences. The probability that consumer i chooses news outlet j on day t on the consumption occasion τ is

$$\pi(y_{it\tau} = j|\theta_i) = \frac{\exp(u_{ijt\tau}(\theta_i))}{1 + \sum_{j'} \exp(u_{ij't\tau}(\theta_i))}, \quad (10)$$

implying the likelihood of θ_i observing a sequence of choices y_i of

$$L(\theta_i|y_i) = \prod_t \prod_\tau \prod_j \pi(y_{it\tau} = j|\theta_i)^{I(y_{it\tau}=j)}. \quad (11)$$

We use a normal distribution on the first-stage prior of θ_i , a normal prior over its mean

⁵¹Out of 214,375 news consumers who visit a news article page at least once over the sample period. While they correspond only to 24.5% of news readers in the market, they account for 92.2% of all the news articles read in the data sample period.

and an inverse Wishart prior over the covariance matrix:

$$\begin{aligned}\theta_i &\sim N(\mu, \Sigma), \\ \mu &\sim N(\bar{\mu}, \Sigma \otimes a_\mu^{-1}), \\ \Sigma &\sim IW(\nu_\Sigma, \Psi_\Sigma).^{52}\end{aligned}\tag{12}$$

The flexibility of this specification comes through an unrestricted covariance matrix Σ , which allows for correlations across all outlet fixed effects and other consumer preferences. This flexibility allows us to capture the alternative heterogeneity explanations for changes in the outlet market shares discussed at the end of Section 5.2. However, the cost of this flexibility is that we cannot account for the potential within-day correlations of the error terms across the consumers; as the result, the sampling procedure might underestimate the uncertainty around the posterior point estimates. Our estimation also comes at a high computational cost, making the MCMC hybrid sampling procedure memory- and time-intensive. Online Appendix K provides more details about the sampling procedure.

7 Results

In this section we present and discuss the posterior point estimates of consumer preferences, break them down by potential mechanisms, discuss the implications, and present the counterfactual simulations with different levels of government control of the news market.

7.1 Consumer Preference Estimates

Table 10 reports the distribution of posterior point estimates of consumer preferences from the model defined in equation 9. First, we summarize the distributions of persistent preferences, α_{ij} , by presenting the average α across the types of news outlets. We demean the average $\hat{\alpha}_j$ within the type, $\hat{\alpha}_{\text{type}}$, by the average α_j across all the news outlets, $\hat{\alpha}$, to make the magnitudes of the estimates more comparable. The estimates reveal that an average consumer has the highest persistent preference for the GC news outlets ($\hat{E}(\hat{\alpha}_{GC} - \hat{\alpha}) = 1.1033$), followed by the independent ($\hat{E}(\hat{\alpha}_{Ind} - \hat{\alpha}) = 0.129$) and potentially influenced ($\hat{E}(\hat{\alpha}_{Inf} - \hat{\alpha}) = 0.128$) news outlets. There is substantial heterogeneity in consumer preferences – for instance, the standard deviation of preferences for the independent outlets, $\hat{\alpha}_{Ind} - \hat{\alpha}$, is 0.592 – meaning that there are a lot of people who prefer an average outlet to the independent news outlets. At the same time, the vast majority (97.5%) of consumers have higher persistent preferences for the GC news outlets than for an average news outlet in this market –

indicating a strong fixed taste for the GC outlets.

Table 10: Posterior point estimates of consumer preferences.

	Mean	S.D.	% of users > 0
$\hat{\alpha}$	-5.872 (0.016)	1.101 (0.012)	0 -
$\hat{\alpha}_{GC} - \hat{\alpha}$	1.103 (0.016)	0.547 (0.01)	97.5 (0.24)
$\hat{\alpha}_{Inf} - \hat{\alpha}$	0.128 (0.008)	0.273 (0.005)	68.65 (1.03)
$\hat{\alpha}_{Ind} - \hat{\alpha}$	0.129 (0.015)	0.592 (0.01)	58.89 (1.07)
$\hat{\alpha}_{Int} - \hat{\alpha}$	-2.253 (0.096)	1.015 (0.048)	1.67 (0.25)
$\hat{\alpha}_{Ukr} - \hat{\alpha}$	-2.532 (0.05)	2.542 (0.034)	14.77 (0.35)
$\hat{\beta}^{POC}$	0.028 (0.002)	0.146 (0.002)	58.85 (0.7)
$\hat{\beta}^{Ukr}$	0.094 (0.003)	0.218 (0.002)	67.2 (0.56)
$\hat{\gamma}$	0.016 (0.002)	0.133 (0.002)	54.98 (0.75)
$\hat{\rho}$	-0.052 (0.004)	0.182 (0.003)	39.9 (0.86)

The posterior standard deviation estimates are in parentheses.

The results are drastically different when we examine preferences of consumers for the news coverage and ideological positions of the outlets. An average consumer prefers POC ($\hat{E}(\hat{\beta}^{POC}) = 0.028$) and Ukraine-crisis ($\hat{E}(\hat{\beta}^{Ukr}) = 0.094$) news to the non-sensitive news, and a more anti-government slant in the Ukraine-crisis news ($\hat{E}(\hat{\gamma}) = 0.016$). This implies that an average consumer has a distaste for the censorship (report less POC news) and the ideological framing (more pro-government slant) of the GC news outlets. Such preferences hold for the majority (58.85% and 54.98%) of consumers in the online news market in Russia.

A negative estimate of the mean of ρ coefficient implies that an average consumer does not read the Ukraine-crisis news like a conscientious type, who would be more likely to sample alternative ideological positions on days with more Ukraine-crisis news. Only a small share of consumers (39.9%) exhibit this type of behavior. Consumers with a high average $\hat{\alpha}$ and those who have a higher preference for the independent and international outlets tend to have higher ρ estimates.

Table 11: Decomposed utility differences between the GC and independent outlets.

	Mean	S.D.	% of users > 0
<i>Persistent preferences:</i>			
$\hat{\alpha}_{GC} - \hat{\alpha}_{Ind}$	0.974 (0.015)	0.851 (0.015)	87.95 (0.54)
<i>Preferences for coverage:</i>			
On days with average volume of sensitive news (average V_t^{POC} and V_t^{Ukr}):			
$\hat{\beta}^{POC}(\bar{x}_{GC}^{POC} - \bar{x}_{Ind}^{POC})$	-0.07 (0.006)	0.369 (0.005)	41.15 (0.7)
$\hat{\beta}^{Ukr}(\bar{x}_{GC}^{Ukr} - \bar{x}_{Ind}^{Ukr})$	0.017 (0)	0.039 (0)	67.2 (0.56)
$\hat{\gamma}(sl_{GC} - sl_{Ind})$	-0.046 (0.007)	0.393 (0.005)	45.02 (0.75)
On days with a lot of sensitive news (V_t^{POC} and V_t^{Ukr} 2 s.d. above average):			
$\hat{\beta}^{POC}(\bar{x}_{GC}^{POC} - \bar{x}_{Ind}^{POC})$	-0.13 (0.011)	0.685 (0.009)	41.15 (0.7)
$\hat{\beta}^{Ukr}(\bar{x}_{GC}^{Ukr} - \bar{x}_{Ind}^{Ukr})$	0.031 (0.001)	0.072 (0.001)	67.2 (0.56)
$\hat{\gamma}(sl_{GC} - sl_{Ind})$	-0.085 (0.013)	0.728 (0.009)	45.02 (0.75)

The posterior standard deviation estimates are in parentheses. \bar{x}_{GC} and \bar{x}_{Ind} represent average reporting positions of the GC and independent outlets, respectively. sl_{GC} and sl_{Ind} represent average ideological framing positions of the GC and independent outlets in the Ukraine-crisis news.

Table 11 uses the estimates to compare the utilities that consumers get from the GC and independent outlets. We decompose the differences into two parts; the part driven by the persistent preferences of the readers, and the part driven by the news outlets' coverage.

Estimates of the persistent preferences reveals that – in the absence of differences in sensitive news coverage – an average consumer gets substantially higher utility from the GC outlet. The mean utility difference is 0.974, with 87.95% of consumers having a higher persistent preference for the GC outlets. At the same time, an average consumer prefers the ideological position of the independent news outlets, even though the magnitude of the utility differences stemming from these preferences is lower than the utility differences from the persistent preferences. Rows 2-4 present the differences in the utility consumers get from

the coverage of the GC and independent outlets on a day with an average volume of sensitive (POC and Ukraine-crisis) news. An average consumer gets only 0.07 extra utils from the POC news coverage on the independent outlets, and only 0.046 extra utils from the less pro-government ideological slant of the independent outlets in the Ukraine-crisis coverage. Since the GC outlets cover the Ukraine crisis slightly more than independent outlets, consumers get 0.017 extra utils from the GC outlets' coverage since they prefer more news about the Ukraine crisis.

Rows 5-7 extend a similar comparison to days with a lot – 2 standard deviations above average – of POC and Ukraine-crisis news. While the magnitudes of the utility differences on such days are higher, utilities consumers get from sensitive news coverage is still lower than the difference driven by the persistent preferences. We conclude that persistent preferences is the primary driver of consumption of the GC news outlets.

Table 12 converts the utility differences between the outlets into the implied market shares under different volumes of POC and Ukraine-crisis news. Column (1) presents the predicted market shares on days with only non-sensitive news, $V_t^{POC} = 0$ and $V_t^{Ukr} = 0$. On such days, controlled outlets are expected to get 14.33% of the market, while the independent outlets are getting only 10.79%. This difference stems from the persistent preferences of the consumers. The implied market share ratio is 1.33, presented in the lower part of the table. As the volume of the POC news (V_t^{POC}) increases, the market share of the independent outlets starts to increase faster than the market share of the GC outlets, reflecting consumers' preference for more coverage of POC news. As a result, on days with an average volume of POC news, the ratio of GC to independent outlets' market share is 1.26, and it goes down to 1.15 on days with a lot (2 standard deviations above the mean) of POC news.

Similarly, columns (4) and (5) of Table 12 present changes in market shares on days with more Ukraine-crisis news. The market share of GC outlets grows slightly faster than the market share of the independent outlets, due to their higher coverage of the Ukraine-crisis news. However, the market shares of the international and Ukrainian outlets grow much faster, with the implied market shares ratio changing from 16.16 to 10.2 for $share_{Gov}/share_{Int}$ and from 10.82 to 5.3 for $share_{Gov}/share_{Ukr}$ on days with a lot (2 standard deviations above the mean) of Ukraine-crisis news. Such differences are driven by consumers' preferences for the anti-government slant in the news.

Figure 10 presents the histogram of differences in consumers' probabilities to choose one of the GC and independent outlets, conditional on choosing one or another, $\frac{Pr(j \in GC) - Pr(j \in Ind)}{Pr(j \in GC) + Pr(j \in Ind)}$. The red histogram corresponds to days with no sensitive news. On such days, the conditional

Table 12: Simulated market shares for different levels of POC and Ukraine-crisis news.

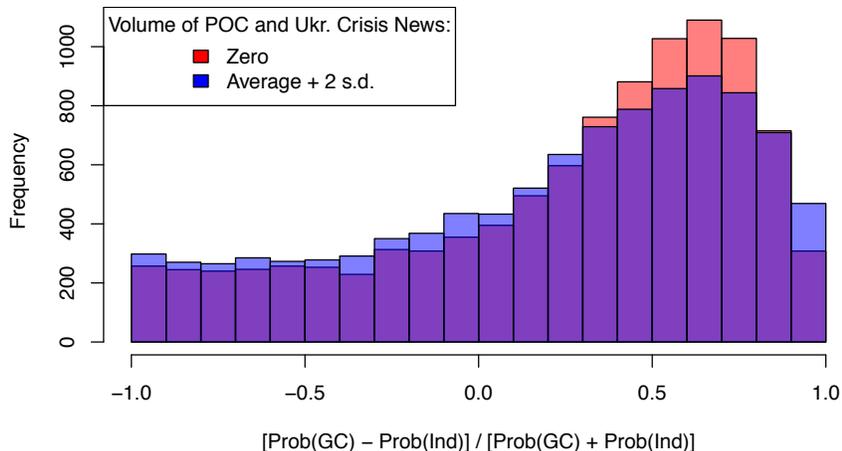
Outlet Types	Market Shares				
	(1)	(2)	(3)	(4)	(5)
	Volume of Sensitive News				
V_t^{POC} :	0	Mean	Mean + 2 S.D.	0	0
V_t^{Ukr} :	0	0	0	Mean	Mean + 2 S.D.
share _{Gov}	14.33 (0.08)	14.58 (0.08)	14.67 (0.08)	15.27 (0.06)	16.64 (0.1)
share _{Inf}	16.11 (0.05)	16.37 (0.06)	16.53 (0.07)	17.23 (0.04)	17.79 (0.05)
share _{Ind}	10.79 (0.06)	11.57 (0.06)	12.72 (0.07)	11.71 (0.05)	12.27 (0.06)
share _{Int}	0.89 (0.01)	0.93 (0.01)	0.98 (0.02)	1.11 (0.01)	1.63 (0.02)
share _{Ukr}	1.32 (0.03)	1.32 (0.03)	1.29 (0.03)	1.68 (0.01)	3.14 (0.03)
share _{Outside}	56.57 (0.14)	55.23 (0.14)	53.8 (0.18)	53 (0.1)	48.53 (0.15)
Market Share Ratios:					
share _{Gov} /share _{Inf}	0.89 (0.01)	0.89 (0.01)	0.89 (0.01)	0.89 (0)	0.94 (0.01)
share _{Gov} /share _{Ind}	1.33 (0.01)	1.26 (0.01)	1.15 (0.01)	1.3 (0.01)	1.36 (0.01)
share _{Gov} /share _{Int}	16.16 (0.23)	15.66 (0.25)	14.9 (0.28)	13.72 (0.14)	10.2 (0.14)
share _{Gov} /share _{Ukr}	10.82 (0.22)	11.09 (0.23)	11.34 (0.24)	9.09 (0.06)	5.3 (0.07)

The market shares are percentages of the entire market. The posterior standard deviation estimates are in parentheses.

probability of choosing a GC outlet for an average consumer is 62.9%, and 37.1% for the independent outlet, meaning the average difference is 25.8 percentage points. Around 72.9% of consumers are more likely to choose a GC outlet. On days with a lot – 2 standard deviations above the average – of POC and Ukraine-crisis news (blue histogram), the conditional probability of choosing a GC outlet for an average consumer reduces to 60.9%, meaning that the average difference with the independent outlets reduces to 21.8 percentage points. While the probability to choose an independent outlet over a GC outlet increases for most consumers on days with a lot of sensitive news, some consumers are more likely to navigate

to the GC outlets on such days – as indicated by a small blue spike on the right side of the histogram. This implies that choices of consumers become more polarized by outlet type on days with more sensitive news.

Figure 10: Differences in choice probabilities of GC and independent outlets, conditional on choosing one or another.



Red histogram corresponds to days with no sensitive news, and blue histogram – to days with 2 standard deviations above the average volume of sensitive news. Histograms are computed for a random MCMC draw – changing the draw does not affect the qualitative results.

7.1.1 The Nature of Persistent Preferences

We now explore the nature of high persistent preferences of consumers for the GC outlets.

First, we characterize the consumption patterns of news readers with a high persistent preference for the GC over independent outlets. For this, we regress the difference in persistent preferences, $\hat{\alpha}_{GC} - \hat{\alpha}_{Ind}$, on how consumers get to the GC outlets’ websites, what type of pages they land on and what type of articles they read there. Table 13 presents the results. Consumers with a high preference for GC outlets over independent outlets are less likely to land on GC outlets through Yandex (row 1), more likely to directly visit a GC outlet (row 2), more likely to land on the main page of the GC outlet (row 4), less likely to read POC news (row 5) and have a lower preference for news outlets in general (row 7). Thus, a typical loyal consumer of GC outlets will navigate to the website directly, through the main page, and will be more likely to read either non-sensitive or Ukraine-crisis news.

Table 13: Correlation of persistent preferences of consumers with the news consumption patterns on the GC outlets.

Share of GC Page Views	<i>Dependent variable:</i>					
	$\hat{\alpha}_{GC} - \hat{\alpha}_{Ind}$					
	(1)	(2)	(3)	(4)	(5)	(6)
First Visit from Yandex	-0.163*** (0.020)					-0.102*** (0.028)
First Visit Direct		0.172*** (0.021)				0.097*** (0.026)
First Land on Other Pages			0.062** (0.025)			-0.037 (0.030)
First Land on Main Page			0.226*** (0.033)			0.129*** (0.034)
Share of POC News Read				-0.284*** (0.042)		-0.283*** (0.042)
Share of Ukr News Read				-0.027 (0.023)		-0.044* (0.023)
$\hat{\alpha}_i$					-0.029*** (0.006)	-0.023*** (0.007)
Constant	1.043*** (0.008)	0.917*** (0.012)	0.964*** (0.010)	1.042*** (0.011)	0.829*** (0.038)	0.896*** (0.047)
Observations	9,579					
R ²	0.007	0.007	0.005	0.005	0.002	0.017
Adjusted R ²	0.006	0.007	0.005	0.005	0.002	0.017

*p<0.1; **p<0.05; ***p<0.01

A news consumer is a unit of observation. The first four variables are shares of GC outlets' page views after the corresponding type of arrival on the website. Variables five and six are shares of news articles read about POC and Ukraine-crisis news, respectively, out of any news articles read.

To get a better understanding on what kind of news are preferred by the loyal customers of the GC outlets, we correlate texts of articles read by consumers with their relative preference for the GC outlets. For each news reader in our estimation sample, we construct a vector of counts of the 2,000 most common bigrams of named entities that this person is exposed to by visiting GC outlets’ news articles. We then perform a penalized Lasso regression (Tibshirani, 1996) of $\hat{\alpha}_{GC} - \hat{\alpha}_{Ind}$ on a matrix of these word counts. Tables A29-A31 in Online Appendix N presents words selected by Lasso as the most correlated with $\hat{\alpha}_{GC} - \hat{\alpha}_{Ind}$.⁵³ News consumers with a high persistent preference for GC over independent outlets tend to read news about celebrities (for example, “Jeanna Friske,” “George Clooney,” “Hollywood Reporter,” “Kate Middleton”), sports (“AK Bars,” “Victor Tikhonov,” “[Paris] Saint-Germain,” “Anastasia Pavlyuchenkova,” “Vladimir Granat”) and international events (“Sierra Leone,” “Saint-Nazaire,” “Japan Shinzo [Abe]”), while consumers with a low preference for the GC outlets are more likely to read articles related to Russian and Ukrainian politics and law (“Western Ukraine,” “Stepan Bandera,” “Pavel Krashennikov,” “Vitaly Klichko,” “Oleh Tyahnybok,” “Alexander Efremov”).

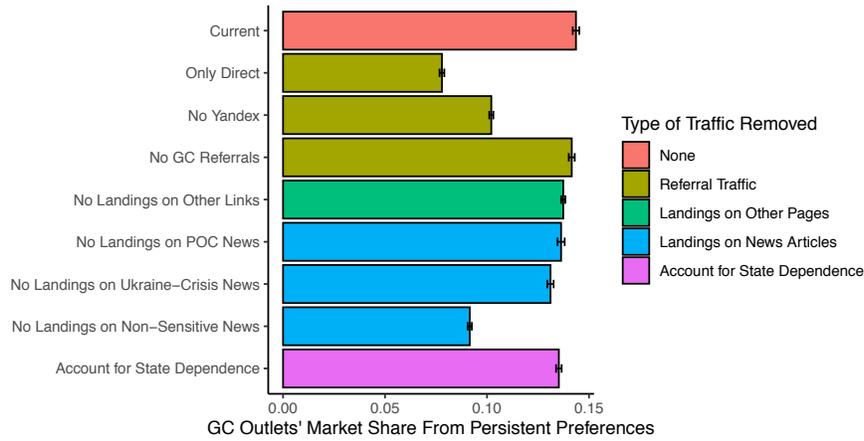
While results in Table 13 show that the most loyal consumers of GC outlets choose to navigate directly and land on outlets’ main pages, a large fraction of GC outlets’ traffic comes from third parties – Yandex in particular – and lands on news articles and pages with video content, as we have seen in Section 5.1. Such traffic generators might play an important role in forming high persistent preferences for GC outlets. In our data, we do not have a natural experiment that exogenously removes these different traffic generators to assess their causal impact; instead, we provide descriptive evidence of the relative importance of each of these mechanisms. For this, we exclude consumption sessions of GC outlets that were started by different types of referral websites and landing pages from the data, and re-estimate the structural model of demand as if GC outlets did not get that traffic.

Figure 11 summarizes the relative importance of different traffic generating mechanisms by presenting simulated market shares of GC outlets driven by the persistent preferences of consumers.⁵⁴ Under the current persistent preferences, GC outlets get 14.4% of the market on days with no sensitive news. If we remove all indirect traffic of GC outlets, this market share would decrease to 7.8%, a 6.6 percentage points reduction. Yandex traffic accounts

⁵³We exclude 20 bigrams that predominantly – more than 50% of the counts – appear on the same outlet; usually these are outlet-specific bigrams of named entities, such as names of journalists or headers under pictures and videos.

⁵⁴Tables A20-A26 in Online Appendix L presents model estimates with different excluded traffic of the GC outlets.

Figure 11: Simulated market shares of GC outlets based only on persistent preferences.



Each bar represents the estimation results with different GC outlet arrivals excluded. We simulate the market shares for days with no sensitive news, $V_t^{POC} = V_t^{Ukr} = 0$, meaning that market shares are solely driven by the persistent preferences of consumers. Error bars correspond to two standard deviations of the MCMC draws.

for the most – 4.2 percentage points – of this reduction, and referrals of other GC outlets – only for 0.2 percentage points. Landings on other pages – including video content – increase the persistent preferences and the corresponding market share of GC outlets by 0.7 percentage points, and landings on POC and Ukraine-crisis – by 0.8 and 1.3 percentage points, respectively. Landings on non-sensitive news play a larger role – without them, persistent preferences of consumers for GC outlets would generate only 9.2% market share, 5.2 percentage points less than currently.

We also examine the degree to which persistent preferences of consumers for GC outlets stem from choice inertia, an accumulated habit of consumers to revisit the same outlet. To capture the choice inertia, we add a state dependence variable to the utility model – an indicator variable taking a value “1” if this GC outlet was visited on the previous day with any news consumption – a common first-order Markov formulation used in the literature measuring brand loyalty (Dubé et al., 2010; Bronnenberg and Dubé, 2017). Online Appendix M writes out the model specification and presents the estimation results.⁵⁵ After excluding the accumulated brand loyalty, persistent preferences of consumers generate a 13.5% market share for GC outlets, a 0.7 percentage points reduction from the current regime.

⁵⁵We handle the initial conditions problem (Heckman, 1981) by estimating the bounds on the state dependence coefficient as proposed by Simonov et al. (2019). The difference in the upper and lower bounds on the state dependence estimate – presented in the last row of Tables A27 and A28 in Online Appendix M.2 – is statistically insignificant, showing that our setting does not suffer from the initial conditions problem.

Finally, persistent preferences of consumers might be driven by fixed characteristics of news outlets, such as the overall quality of the website and long-term effects of the outlets’ ideology. Once again, we do not observe exogenous changes in such characteristics; instead, we describe their relative importance in α_{ij} by exploiting an estimated correlation in persistent preferences across outlets. First, for each consumer i , we demean α_{ij} by $\bar{\alpha}_i$, to exclude the overall preference of this consumer for visiting news outlets. We then compute correlations in $\alpha_{ij} - \bar{\alpha}_i$ for each pair of outlets $j, j' : j \neq j'$, by using posterior point estimates of α_{ij} for consumers in our sample. This provides us with $42 \cdot 42 / 2 - 42 / 2 = 861$ unique correlation estimates. Figure A22 in Online Appendix O visualizes these estimates.

To test whether news outlets with similar characteristics also have more similar persistent preference, we regress the correlation estimates of news outlet pairs, $\hat{c}(\alpha_{ij} - \bar{\alpha}_i, \alpha_{ij'} - \bar{\alpha}_i) : j \neq j'$, on the absolute value of the difference in outlets’ characteristics z , $|z_j - z_{j'}|$. Table 14 presents the estimates. Two outlets are more likely to be preferred by the same consumer if they have a more similar average persistent preference across consumers, $\bar{\alpha}_j$ (a proxy for the outlet’s quality), more similar reporting and ideological positions, sl_j , \bar{x}_j^{POC} and \bar{x}_j^{Ukr} , and if a more similar share of their traffic is referred by Yandex. All absolute distance variables are normalized to have a unit standard deviation, so more negative coefficients represent a stronger correlation between a preference and characteristic similarity. In particular, ideological framing similarity of outlets is the strongest in predicting outlets’ preference similarity; outlets with one standard deviation more similar ideological framing have a 0.046 higher correlation in the persistent preferences. This result holds even after we control for the referral traffic from Yandex and for outlets being of the same type – suggesting that outlets’ ideological positions enter the persistent preferences of consumers and have a long-term impact.

7.2 Counterfactuals

Consumer preference estimates have revealed that persistent preferences play an important role in consumer demand for GC outlets – an average consumer has a strong preference for GC outlets but prefers the average reporting and ideological position of the independent outlets. We now assess the degree to which GC outlets benefit from strong persistent tastes of consumers, as well as what is the “cost” of the potentially sub-optimal ideological positions of the GC and potentially-influenced outlets.

To understand the impact of government control on the outlets’ market shares, we simulate market outcomes in counterfactual scenarios with different ideological positions of news

Table 14: Relationship between the correlations in persistent preferences of consumers, $\alpha_{ij} - \bar{\alpha}_i$, and distance between the outlets' characteristics.

	<i>Dependent variable:</i>					
	$\text{c\hat{o}r}(\alpha_{ij} - \bar{\alpha}_i, \alpha_{ij'} - \bar{\alpha}_i) \forall j \neq j'$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-0.030** (0.010)	0.052*** (0.016)	0.105*** (0.024)	0.014 (0.012)	0.150*** (0.025)	0.148*** (0.023)
$ \frac{\# \text{ articles}}{\text{day}}_j - \frac{\# \text{ articles}}{\text{day}}_{j'} $	0.006 (0.006)				0.012* (0.009)	0.012* (0.009)
$ \frac{\# \text{ words}}{\text{article}}_j - \frac{\# \text{ words}}{\text{article}}_{j'} $	0.008 (0.007)				0.007 (0.008)	0.009 (0.007)
$ \bar{\alpha}_j - \bar{\alpha}_{j'} $		-0.056*** (0.011)			-0.029*** (0.011)	-0.027*** (0.01)
$ \text{sl}_j - \text{sl}_{j'} $			-0.051*** (0.009)		-0.046*** (0.009)	-0.047*** (0.009)
$ \bar{x}_j^{POC} - \bar{x}_{j'}^{POC} $			-0.021*** (0.007)		-0.024*** (0.008)	-0.022*** (0.008)
$ \bar{x}_j^{Ukr} - \bar{x}_{j'}^{Ukr} $			-0.036*** (0.009)		-0.029*** (0.010)	-0.030*** (0.009)
$ \% \text{ Yand.}_j - \% \text{ Yand.}_{j'} $				-0.027** (0.013)	-0.038*** (0.014)	-0.039*** (0.013)
$j, j' \in \text{GC}$						0.064** (0.028)
$j, j' \in \text{Ukr}$						0.740*** (0.017)
$j, j' \in \text{Ind}$						0.007 (0.031)
$j, j' \in \text{Inf}$						-0.044* (0.024)
$j, j' \in \text{Int}$						0.197*** (0.023)
Observations			861			
R ²	0.002	0.074	0.127	0.018	0.179	0.241
Adjusted R ²	-0.00002	0.073	0.124	0.017	0.172	0.231

*p<0.1; **p<0.05; ***p<0.01

Standard errors are clustered at the outlet level. All absolute distance variables are normalized to have a unit standard deviation. The last five variables are indicators of outlets belonging to the same outlet type. Variable % Yand._j refers to a share of page views generated after the first arrival to the website from Yandex.

outlets. The government controls its outlets with censorship – which decreases the share of news reporting about POC news (low $\bar{x}_j^{POC} \forall j \in GC$) – and ideological framing in the Ukraine-crisis news (low $sl_j \forall j \in GC$). To simulate market shares without government control, we adjust the share of reporting of GC news outlets about POC news and their ideological framing of Ukraine-crisis news so that the average values for GC outlets match the independent outlets. More specifically, we adjust $\bar{x}_j^{POC*} = \bar{x}_j^{POC} * (\bar{x}_{Ind}^{POC} / \bar{x}_{GC}^{POC})$ and $sl_j^* = sl_j - sl_{GC} + sl_{Ind}$ for all $j \in GC$, where \bar{x}_{GC}^{POC} and \bar{x}_{Ind}^{POC} represent average reporting positions of the GC and independent outlets about POC news, and sl_{GC} and sl_{Ind} represent average ideological framing positions of the GC and independent outlets in the Ukraine-crisis news. By doing this, we treat average ideological positions of the independent news outlets as “unbiased.” We interpret simulation results as short-term reactions of the market to changes in the level of government control.⁵⁶

We simulate the market shares with new \bar{x}_j^{POC*} and sl_j^* for different realizations of sensitive news, V_t^{POC} and V_t^{Ukr} , and report the market shares averaged over time in Table 15.⁵⁷ Column (2) reports the predicted market shares with adjusted \bar{x}_j^{POC*} and sl_j^* for the GC outlets – a case when government does not exercise direct control of the news market through ownership (Gehlbach and Sonin, 2014). The market share of GC outlets increases from the current 15.56% to 17.94%, a 2.38 percentage points (15.3%) increase. More than half (1.13 p.p.) of this increase is coming from the outside option (extensive margin), and the rest is mainly covered by the potentially-influenced and independent outlets.

Similarly, in column (3) we compute the “cost” of government control for the potentially-influenced outlets – we adjust their average POC news reporting and ideological framing in Ukraine-crisis news to match independent outlets. The potentially-influenced outlets are not owned but still partially controlled by the government, representing indirect control (Gehlbach and Sonin, 2014). If they were to report like the independent outlets, their expected market share would increase by 1.49 percentage points to 18.92%, an 8.5% increase to the current expected market share.

Column (4) simulates the market under no direct and indirect control and confirms the above results, although in this case the market shares of the GC and potentially influenced

⁵⁶In the long run, we would expect changes both on the supply side, such as product differentiation decisions, and on the demand side, such as changes in persistent preferences. Further, when changing the reporting and ideological positions of the GC outlets, we assume that they retain their persistent preferences, which in part might be driven by the high quality of their non-sensitive news coverage.

⁵⁷In order to speed up the counterfactual simulation, we approximate news realizations V_t^{POC} and V_t^{Ukr} by the centers of 20 clusters of these variables and simulate one choice occasion per consumer per day. Standard k-means clustering algorithm is applied to cluster the observed V_t^{POC} and V_t^{Ukr} .

Table 15: Simulated market shares for different levels of government control and persistent preferences for the GC news outlets.

Outlet Types	Market Shares					
	(1)	(2)	(3)	(4)	(5)	(6)
	Level Of Governments' Control					
	Actual	No control		More control		Low
		Direct	Indirect	Both		α_{GC}
share _{Gov}	15.56 (0.04)	17.94 (0.12)	15.23 (0.04)	17.33 (0.11)	15.72 (0.04)	7.11 (0.02)
share _{Inf}	17.43 (0.04)	16.82 (0.04)	18.92 (0.08)	18 (0.05)	17.64 (0.04)	19.68 (0.05)
share _{Ind}	12.53 (0.03)	12.01 (0.04)	12.17 (0.04)	11.79 (0.05)	11.73 (0.06)	13.63 (0.04)
share _{Int}	1.2 (0.01)	1.13 (0.01)	1.15 (0.01)	1.09 (0.01)	1.22 (0.01)	1.32 (0.01)
share _{Ukr}	1.8 (0.02)	1.76 (0.02)	1.78 (0.02)	1.75 (0.02)	1.81 (0.02)	1.92 (0.02)
share _{Outside}	51.47 (0.06)	50.34 (0.1)	50.76 (0.07)	50.03 (0.12)	51.88 (0.07)	56.34 (0.07)

The market shares are percentages of the entire market. The posterior standard deviation estimates are in parentheses.

outlets increase slightly less compared to the current benchmark – which is intuitive given that similar (unbiased) ideological positions of all outlets intensify competition.

In column (5) we examine the reverse scenario of more indirect control, a case when the independent news outlets change their ideological positions to the ones of the potentially influenced outlets.⁵⁸ In this case, the market share of independent news decreases from 12.53% to 11.73%, a 0.8 percentage points reduction.

Simulations show that news outlets lose from 0.8 to 2.38 percentage points of the market share due to the government control. To assess the amount of money at stake, we do a simple back-of-the-envelope calculation. Almost all of the online news outlets in Russia do not have a paywall, meaning that display advertising is the primary source of their revenue. In 2014, the total expenditure on display advertising on the Russian internet was 19.1 billion rubles

⁵⁸This is perhaps a more feasible scenario given the events of 2016-2017 – by the middle of 2016, several independent news outlets had to change their ownership due to a new law (TrustLaw, 2016), and rbc, one of the top online news outlets in Russia, had to change the editorial team due to the government pressure (bbc.com, 2016) as well as its ownership later in 2017 (forbes.ru, 2017).

(akarussia.ru, 2014), or around \$318 million using a 60 rubles for a dollar exchange rate (exchange rates.org, 2014). If we assume that the online news market gets all the display advertising revenues – a generous best-case scenario for the news outlets – 1 percentage point of the news market share converts to $\$318 * 0.01 / (1-0.515) = \6.56 million of display advertising revenue. This implies that GC outlets lose at most $2.38 * \$6.56 = \15.6 million of display advertising revenue per year due to government control, and independent outlets would lose $0.8 * \$6.56 = \5.25 million if they became controlled. For comparison, government subsidies to mass media in Russia in 2015 were \$1.21 billion (rbc.ru, 2015) – several orders of magnitude more than the potential loss of the online outlets.

Finally, in column (6) of Table 15 we present the expected market shares of news outlets with lower average persistent preferences of GC news outlets. For each consumer i , we adjust the level of GC outlet preferences so that the average preference of consumers for GC outlets matches the average persistent preference of the independent outlets: $\alpha_{ij}^* = \alpha_{ij} - \hat{\alpha}_{GC} + \hat{\alpha}_{Ind} \forall j \in GC$. Under the lower persistent preference regime, the market share of the GC news outlets decreases by 8.45 percentage points, or 54.3% – meaning that high persistent preferences of the GC outlets is around $8.45/2.38 \approx 3.5$ times more important in generating their market share than removing the government control of the news. While we cannot causally separate out the source of high persistent preference of the GC outlets, their description in Section 7.1.1 suggests that such high increase in market share is driven by a high referral traffic of GC outlets and their coverage of non-sensitive topics – for instance, news about celebrities and sports.

7.2.1 Online Media Power of the Government

While market shares and the corresponding display advertising revenues are important for the GC news outlets, the main reason for government’s investments into the GC outlets is to capture the attention of the news readers and potentially persuade them to support the government. To understand the ability of the government to influence readers in the online news market in Russia, we compute the degree of media power (Prat, 2017) that the GC outlets have, as well as the role of high persistent preferences in this media power. Given that we do not have access to cross-platform news consumption data like Kennedy and Prat (2017), we focus solely on the online news market and compute the degree of online media power.

First, we extend the definition of the attention share in Prat (2017) to our model set-up.

The attention share of consumer i on day t to an outlet j is

$$\text{attention share}_{ijt} = \frac{\Pr(y_{it} = j)}{1 - \Pr(y_{it} = 0)}, \quad (13)$$

where 0 is an outside option of not reading the news. Aggregating this across days and consumers, we get the overall attention share of an outlet j

$$\text{attention share}_j = \sum_{t=1}^T \sum_{i=1}^I \frac{\text{attention share}_{ijt}}{I * T}. \quad (14)$$

The attention share of the GC news outlets is then

$$\text{attention share}_{GC} = \sum_{j \in GC} \sum_{t=1}^T \sum_{i=1}^I \frac{\text{attention share}_{ijt}}{I * T}. \quad (15)$$

Table 16: Attention shares and market power of GC outlets under alternative persistent preferences

	Alternative Persistent Preferences for GC Outlets							
	(1) Current	(2) Low α (Like indep.)	(3) No Referrals Only Direct	(4) No Yandex	(5) No Other Links	(6) No Article Arrivals POC	(7) Ukraine Crisis	(8) Not sens.
att. share _{GC}	33.8 (0.08)	17.92 (0.05)	18.51 (0.08)	24.29 (0.06)	32.72 (0.07)	32.13 (0.08)	28.61 (0.07)	24.71 (0.06)
Media power _{GC}	0.511 (0.002)	0.218 (0.001)	0.227 (0.001)	0.321 (0.001)	0.486 (0.001)	0.473 (0.002)	0.401 (0.001)	0.328 (0.001)

Column (1) of Table 16 presents the attention share estimates of GC outlets. Under the current persistent preferences of consumers, GC outlets capture 33.8% of online news consumers’ attention. This attention share corresponds to the upper bound of 0.51 on governments’ media power, meaning that the government is able to swing 24.5-75.5% elections into a draw.⁵⁹

Column (2) of Table 16 presents GC outlets’ attention shares and media power under a lower level of persistent preferences of consumers – as if the average preference of consumers for GC outlets was the same as for independent outlets (similar to the simulation in column (6) of Table 15). In this case, the attention share of GC outlets would be 17.92%, a 15.88

⁵⁹The upper bound is computed based on the “worst-case scenario” assumptions – that readers are naive and do not understand that the GC news outlets are trying to persuade them (Prat, 2017).

percentage points reduction. In this case, media power of the government would be only 0.218, meaning it can swing only 39-61% elections into a draw.

Columns (3)-(8) of Table 16 provide descriptive evidence on the relative importance of alternative mechanisms behind the high persistent preferences for GC outlets. For this, we exclude consumption sessions of GC outlets that were started by different types of referral websites and landing pages from the data, and then re-estimate the structural model of demand as if GC outlets did not get that traffic.⁶⁰ We find that indirect traffic (column 3) – and Yandex in particular (column 4) – play a very important role in increasing the media power of the government; attention share of GC outlets would be 18.5% if the persistent preferences did not benefit from the indirect traffic, and 24.3% if they did not benefit from the traffic from Yandex. Availability of other pages, POC news and Ukraine-crisis news slightly improve persistent preferences of consumers for GC outlets, increasing their attention share by 1.1 (column 5), 1.7 (column 6) and 5.2 (column 7) percentage points, respectively. The increase in persistent preferences from the availability of non-sensitive news increases the attention share of the government by 9.1 percentage points (column 8).

Finally, demand estimates allow us to examine the degree to which GC outlets can capture the attention of consumers who prefer the news coverage of independent outlets – a group of consumers that is more likely to be opposing the incumbent government in voting. Capturing the attention of these consumers is particularly important on days with a lot of sensitive news events – since the government does not want them to be exposed to sensitive news. The GC outlets have an attention share of 31.5% among the consumers who prefer more POC news coverage ($\hat{\beta}_i^{POC} > 0$) and on days with a lot – 2 standard deviations above average – of POC news. A 15.15 percentage points of this attention share are driven by the high persistent preferences of these consumers for the GC outlets. Similarly, on days with a lot of Ukraine-crisis news and among consumers who prefer the anti-government ideological framing in Ukraine-crisis news ($\hat{\gamma}_i > 0$), the attention share of GC outlets is 29.2%, with 14.3 percentage points driven by the high persistent preferences of consumers for the GC outlets.

⁶⁰This is the same set of estimates that we use in Figure 11 – they are presented in Tables A20–A28 in Online Appendix L.

8 Conclusions

In the new era of broad and unrestricted access to information, it is critical to understand whether governments can control public opinion online. In this paper, we show that consumers in the Russian online news market read the GC news outlets even though they have a distaste for the pro-government ideological coverage. Instead, the main source of demand for the GC news outlets comes from the outlet-level tastes of consumers, and data suggests that it is largely driven by third-party referrals and the availability of celebrity news and sports on the GC outlets' websites. Such outlet-level drivers of consumption help the government to impose its sensitive news coverage on the news readers and potentially persuade them to change their ideological preferences.

Our results should be interpreted with two caveats in mind. First, the ideological preferences of consumers in our sample might not extrapolate to the entire population in Russia – we study only online news consumers, whereas TV is still the main news source for an average news consumer in Russia (VTsIOM, 2017). Indeed, most political surveys have indicated the overwhelming support of the government during the period of our study (Economist, 2016), and it is unclear whether our estimates differ because of a bias in the stated preferences in the surveys (Kuechler, 1998) or because of selection on the ideological preferences to consuming news online. However, the ideological preferences of the online news consumers are important on their own – the share of people getting their news online steadily grows, both in Russia (VTsIOM, 2017) and abroad (PewResearchCenter, 2017). Our news consumers also come from the Internet Explorer Toolbar data; these users tend to be older, more work-oriented, and perhaps less technologically-savvy than an average news consumer in Russia. If there is any selection in terms of news preferences of such consumers, demographics suggests (republic.ru, 2012) that the IE Toolbar users should have more pro-government tastes than the average online news reader in Russia – reinforcing our conclusions.

Second, our data and empirical setting does not allow to causally pin down and separate out all alternative mechanisms behind the high persistent tastes of consumers for the GC outlets. While we have presented strong suggestive evidence that highlights the role of third-party referrals and GC outlets' investments in non-sensitive news content, further work is required in this direction. In particular, it is unclear to what extent the current ideological positions of the GC outlets play a role in forming their persistent preferences – and, if there is a long-term effect of ideological positions on news consumption, how long it will take for the persistent preferences to adjust if the ideological positions are changed. Studying this question requires exogenous shocks in the long-term ideological positions of the news outlets,

as well as estimating a model of consumer belief formation – which is an important area for future research.

References

- akarussia.ru (2014). Advertising spending by channel in 2014 in russia [in russian]. http://www.akarussia.ru/knowledge/market_size/id5354 [Accessed: December 16, 2019].
- Allan, J., J. Carbonell, G. Doddington, J. Yamron, Y. Yang, et al. (1998). Topic detection and tracking pilot study: Final report. In *Proceedings of the DARPA broadcast news transcription and understanding workshop*, Volume 1998, pp. 194–218. Citeseer.
- Athey, S., M. M. Mobius, and J. Pál (2017). The impact of aggregators on internet news consumption. *Working Paper*.
- Bai, J., M. Golosov, N. Qian, and Y. Kai (2015). Understanding the influence of government controlled media: Evidence from air pollution in china. *Working Paper*.
- Bartik, T. J. (1991). Who benefits from state and local economic development policies?
- bbc.com (2014). Websites of putin critics blocked in russia. <http://www.bbc.com/news/technology-26578264> [Accessed: December 16, 2019].
- bbc.com (2016). Management of rbc is fired [in russian]. https://www.bbc.com/russian/news/2016/05/160513_rbc_badanin [Accessed: December 16, 2019].
- Besley, T. and A. Prat (2006). Handcuffs for the grabbing hand? media capture and government accountability. *The American Economic Review* 96(3), 720–736.
- Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of machine Learning research* 3(Jan), 993–1022.
- Borusyak, K., P. Hull, and X. Jaravel (2018). Quasi-experimental shift-share research designs. *Working paper*.
- Bronnenberg, B. J. and J.-P. Dubé (2017). The formation of consumer brand preferences. *Annual Review of Economics* 9, 353–382.
- Cagé, J., N. Hervé, and M.-L. Viaud (2019). The production of information in an online world. *The Review of Economic Studies* Forthcoming.

- Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics* 29(2), 238–249.
- Crabtree, C., C. J. Fariss, and H. L. Kern (2015). Truth replaced by silence: A field experiment on private censorship in russia. *Working Paper*.
- Djankov, S., C. McLiesh, T. Nenova, and A. Shleifer (2003). Who owns the media? *The Journal of Law and Economics* 46(2), 341–382.
- Dubé, J.-P., G. J. Hitsch, and P. E. Rossi (2010). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics* 41(3), 417–445.
- Durante, R. and B. Knight (2012). Partisan control, media bias, and viewer responses: Evidence from berlusconi’s italy. *Journal of the European Economic Association* 10(3), 451–481.
- Durante, R., E. V. Zhuravskaya, et al. (2015). *Attack when the World is Not Watching?: International Media and the Israeli-Palestinian Conflict*. Centre for Economic Policy Research.
- Economist, T. (2016). Vladimir putin’s unshakeable popularity. <https://www.economist.com/graphic-detail/2016/02/04/vladimir-putins-unshakeable-popularity> [Accessed: December 16, 2019].
- Edmond, C. (2013). Information manipulation, coordination, and regime change. *Review of Economic Studies* 80(4), 1422–1458.
- Enikolopov, R., M. Petrova, and E. Zhuravskaya (2011). Media and political persuasion: Evidence from russia. *The American Economic Review* 101(7), 3253–3285.
- exchange rates.org (2014). Usd to rur exchange rate on december 31, 2014. <https://www.exchange-rates.org/Rate/USD/RUB/12-31-2014> [Accessed: December 16, 2019].
- forbes.ru (2017). Rbc is sold: Berezkin bought the largest media publisher from prokhorov [in russian]. <https://www.forbes.ru/milliardery/346333-berezkin-kupil-u-prohorova-rbk> [Accessed: December 16, 2019].
- Garcia-Arenas, J. (2016). The impact of free media on regime change: Evidence from russia. *Working Paper*.

- Gehlbach, S. and K. Sonin (2014). Government control of the media. *Journal of Public Economics* 118, 163–171.
- Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *The American Economic Review* 97(3), 713–744.
- Gentzkow, M., B. T. Kelly, and M. Taddy (2017). Text as data. Technical report, National Bureau of Economic Research.
- Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? evidence from us daily newspapers. *Econometrica* 78(1), 35–71.
- Gentzkow, M. and J. M. Shapiro (2015). Ideology and online news. In *Economic Analysis of the Digital Economy*, pp. 169–190. University of Chicago Press.
- Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2011). The effect of newspaper entry and exit on electoral politics. *The American Economic Review* 101(7), 2980–3018.
- Gentzkow, M., J. M. Shapiro, and D. F. Stone (2016). Media bias in the marketplace: Theory. Vol. 2., *Handbook of Media Economics*.
- Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica* 87(4), 1307–1340.
- Groseclose, T. and J. Milyo (2005). A measure of media bias. *The Quarterly Journal of Economics* 120(4), 1191–1237.
- Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process.
- Heckman, J. J. (1991). Identifying the hand of past: Distinguishing state dependence from heterogeneity. *The American Economic Review* 81(2), 75–79.
- Hotelling, H. (1929). Stability in competition. *The economic journal* 39(153), 41–57.
- Hu, L., J. Li, Z. Li, C. Shao, and Z. Li (2013). Incorporating entities in news topic modeling. In *Natural Language Processing and Chinese Computing*, pp. 139–150. Springer.
- Iyengar, S. and D. R. Kinder (1987). News that matters: Agenda-setting and priming in a television age. *University of Chicago Press*.

- Keane, M. P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics* 15(3), 310–327.
- Kennedy, P. and A. Prat (2017). Where do people get their news? *Working Paper*.
- Kim, H., Y. Sun, J. Hockenmaier, and J. Han (2012). Etm: Entity topic models for mining documents associated with entities. In *2012 IEEE 12th International Conference on Data Mining*, pp. 349–358. IEEE.
- Kim, J., G. M. Allenby, and P. E. Rossi (2002). Modeling consumer demand for variety. *Marketing Science* 21(3), 229–250.
- Knight, B. and A. Tribin (2016). The limits of propaganda: Evidence from chavez’s venezuela. *Working Paper, National Bureau of Economic Research*.
- Krasnashchok, K. and S. Jouili (2018). Improving topic quality by promoting named entities in topic modeling. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 247–253.
- Kuechler, M. (1998). The survey method: An indispensable tool for social science research everywhere? *American Behavioral Scientist* 42(2), 178–200.
- Kumaran, G. and J. Allan (2004). Text classification and named entities for new event detection. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 297–304. ACM.
- Kumaran, G. and J. Allan (2005). Using names and topics for new event detection. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 121–128. Association for Computational Linguistics.
- liveinternet.ru (2014). Web page with weekly public ranking of news websites in russia [in russian, accessed through wayback machine]. <https://web.archive.org/web/20141214080407/http://www.liveinternet.ru/rating/ru/media/week.html> [Accessed: January 26, 2020].
- Martin, G. J. and A. Yurukoglu (2017). Bias in cable news: Persuasion and polarization. *American Economic Review Forthcoming*.
- McAlister, L. and E. Pessemier (1982). Variety seeking behavior: An interdisciplinary review. *Journal of Consumer research* 9(3), 311–322.

- McCombs, M. E. and D. L. Shaw (1972). The agenda-setting function of mass media. *Public opinion quarterly* 36(2), 176–187.
- meduza.io (2019). Very bad businessmen: How internet, propaganda and ambitions of konstantin ernst have mad channel one unprofitable – and what will happen to it now [in russian]. <https://meduza.io/feature/2019/12/18/ochen-plohie-biznesmeny> [Accessed: January 26, 2020].
- Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111–3119.
- Mullainathan, S. and A. Shleifer (2005). The market for news. *The American Economic Review* 95(4), 1031–1053.
- Petrova, M. (2008). Inequality and media capture. *Journal of public Economics* 92(1), 183–212.
- PewResearchCenter (2017). Key trends in social and digital news media. <https://www.pewresearch.org/fact-tank/2017/10/04/key-trends-in-social-and-digital-news-media/> [Accessed: December 16, 2019].
- Prat, A. (2017). Media power. *Journal of Political Economy Forthcoming*.
- Prat, A. and D. Strömberg (2013). The political economy of mass media. *Advances in Economics and Econometrics Volume 2, Applied Economics*.
- Qin, B., D. Strömberg, and Y. Wu (2017). Why does china allow freer social media? protests versus surveillance and propaganda. *Journal of Economic Perspectives* 31(1), 117–40.
- Qin, B., D. Strömberg, and Y. Wu (2018). Media bias in china. *American Economic Review* 108(9), 2442–76.
- rbc.ru (2015). Channel one and vgtrk will get compensated for devaluation [in russian]. <https://www.rbc.ru/politics/29/06/2015/55912ffa9a7947453982cda9> [Accessed: December 16, 2019].
- republic.ru (2012). Eight figures: Who are these people supporting putin? [in russian]. <https://republic.ru/posts/1/745894> [Accessed: December 16, 2019].

- Roberts, M. E. (2014). Fear or friction? how censorship slows the spread of information in the digital age. *Working Paper*.
- Rossi, P. (2014). *Bayesian non-and semi-parametric methods and applications*. Princeton University Press.
- Rossi, P. E., G. M. Allenby, and R. McCulloch (2005). *Bayesian statistics and marketing*. John Wiley & Sons.
- Rudolph, M., F. Ruiz, S. Athey, and D. Blei (2017). Structured embedding models for grouped data. In *Advances in Neural Information Processing Systems*, pp. 251–261.
- Schimpfoss, E. and I. Yablokov (2014). Coercion or conformism? censorship and self-censorship among russian media personalities and reporters in the 2010s. *Demokratizatsiya: The Journal of Post-Soviet Democratization* 22(2), 295–311.
- Sen, A. and P. Yildirim (2016). Clicks bias in editorial decisions: How does popularity shape online news coverage? *Working paper*.
- Simonov, A., J.-P. H. Dubé, G. J. Hitsch, and P. E. Rossi (2019). State-dependent demand estimation with initial conditions correction. Technical report, National Bureau of Economic Research.
- squirepattonboggs.com (2014). Russia moves to limit foreign ownership in the media section. http://www.squirepattonboggs.com/~media/files/insights/publications/2014/10/russia-moves-to-limit-foreign-ownership-in-the-2_/files/russiamovestolimitforeignownership/fileattachment/russiamovestolimitforeignownership.pdf [Accessed: December 16, 2019].
- statcounter.com (2015). Desktop browser market share russian federation. <https://gs.statcounter.com/browser-market-share/desktop/russian-federation/> [Accessed: December 16, 2019].
- stopfake.org (2014). A fight with the dishonest information about the ukraine events [in russian]. <http://www.stopfake.org/propaganda-2-0-novye-izmereniya-v-dejstvii/> [Accessed: December 16, 2019].
- Strömberg, D. (2004). Mass media competition, political competition, and public policy. *The Review of Economic Studies* 71(1), 265–284.

- themoscowtimes.com (2013). Kremlin helps media moguls expand. <https://www.themoscowtimes.com/2013/10/20/kremlin-helps-media-moguls-expand-a28748> [Accessed: December 16, 2019].
- themoscowtimes.com (2014). The media war behind the ukraine crisis. <https://www.themoscowtimes.com/2014/03/10/the-media-war-behind-the-ukraine-crisis-a32837> [Accessed: December 16, 2019].
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58(1), 267–288.
- time.com (2014). Russians rewrite history to slur ukraine over war. <https://time.com/3545855/russia-ukraine-war-history/> [Accessed: December 16, 2019].
- Tirole, J. (1988). *The theory of industrial organization*. MIT press.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- TrustLaw (2016). Media regulation in russia: A landscape analysis of laws and trends. <https://www.trust.org/contentAsset/raw-data/4798c68a-eed1-4660-b7c9-fc16a0032cc9/file> [Accessed: December 16, 2019].
- vesti.ru (2014). 116 people burned alive by fascists in odessa [in russian]. <https://www.vesti.ru/doc.html?id=1550135> [Accessed: December 16, 2019].
- VTsIOM (2017). Internet against tv: The battle continues [in russian]. <https://wciom.ru/index.php?id=236&uid=116190> [Accessed: December 16, 2019].
- Xiang, Y. and M. Sarvary (2007). News consumption and media bias. *Marketing Science* 26(5), 611–628.
- Zhu, Y. and A. Dukes (2015). Selective reporting of factual content by commercial media. *Journal of Marketing Research* 52(1), 56–76.