What Drives Demand for Government-Controlled News in Russia?*

Andrey Simonov
Columbia University

Justin Rao
HomeAway, Inc.

8th February 2018

News consumers in many authoritarian countries read government-controlled sources even when independent sources are available. What drives this demand for government-controlled news? We separate out two potential sources of such demand, preferences for pro-government coverage of sensitive events and persistent preferences for news outlets. Identification strategy relies on exogenous shifts in the volume of sensitive events over time. Demand estimates in the Russian online news market reveal that an average consumer has a distaste for pro-government ideology but strong persistent preferences for state-owned outlets. The state-owned outlets would have 20.3% higher market share without the pro-government bias in their reporting.

*Simonov (corresponding author): asimonov@gsb.columbia.edu; Rao: justinmrao@outlook.com. Andrey is grateful to his advisors and dissertation committee members Pradeep Chintagunta, Jean-Pierre Dubé, Matthew Gentzkow and Günter Hitsch for their guidance and support. We thank Charles Angelucci, Ruben Enikolopov, Matt Goldman, Elisabeth Honka, Sanjog Misra, Chris Nosko, 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 and the 15th Media Economics Workshop, seminar participants at the University of Chicago, Stigler Center at Chicago Booth, HKUST, University of Minnesota, University of Rochester, UCSD, Duke, LBS, Columbia, Southern Methodist University, Berkeley, Stanford, CMU, UCLA and University of Southern California for advice and suggestions. All opinions represent our own and not those of our current or past employers. All remaining errors are our own.

1 Introduction

On May 2, 2014, 48 people were killed in violent clashes between the supporters and opposers of the new Ukrainian government in the city of Odessa, southern Ukraine. Fourty-two of them, members of the Ukrainian government opposition, died in the burning of the trade unions building. This was the worst violence to that date in the Ukraine crisis and was 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 have 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. Government-controlled (GC) Russian news outlets had a very different take on the story, reporting 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 outright false information, exemplified by the title of an article of one of the news outlets, vesti.ru: "116 people burned alive by fascists in Odessa."[1]

Every day, news consumers in Russia decide where to read the news. In the online news market in Russia, there is always a choice between 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. Almost none of the Russian news websites have a subscription firewall, so switching from one to another and finding preferred news content is simple. In this environment, a lot of news consumers choose to read the news from the GC outlets and not from the independent outlets, with 4 out of the top 5 online outlets in Russia in late 2014 being either government-owned or potentially influenced.[2]

In this paper, we aim to understand what drives this demand for the GC news outlets in Russia. We distinguish two families of potential explanations. First, consumers might read the GC outlets because of the pro-government bias in the news coverage. There are multiple potential mechanisms behind such interest. Consumers might hold political beliefs that are consistent with the pro-government bias, as suggested by the 80% approval rating of President Putin during the Ukraine conflict,[3] so they prefer the ideological coverage of the

[1] `https://www.vesti.ru/doc.html?id=1550135`

[2] Based on the statistics from `http://www.liveinternet.ru`. Historical data is accessed using the Wayback Machine, `http://web.archive.org/web/20141231225456/http://www.liveinternet.ru/rating/ru/media/week.html`. For the news outlets classification, see Table 1.

[3] Historical ratings of Putin's popularity compiled by The Economist using data from the Levada Center:

GC news outlets because of the confirmation bias (Gentzkow and Shapiro, 2010). Another explanations might be that consumers are "conscientious" and value knowing the government's ideological position about sensitive news events (Mullainathan and Shleifer, 2005; Xiang and Sarvary, 2007) or consumers are simply entertained by the emotional language of the biased coverage (Thussu, 2008). Second, consumers might have a distaste for the GC news outlets' ideological positions but have a strong persistent preference for these news outlets. Such persistent preference can be driven by a number of factors, including a broad and high quality news coverage of the non-sensitive news topics and a convenient and modern website, accumulated brand capital of the outlet, or other sources of product differentiation that are not related to the outlet's ideology.

Separating these potential explanations for the GC outlets' demand is important for two reasons. First, if the GC outlets' demand comes primarily from the persistent preferences of consumers for these outlets, the government has an effective method of control over the ideological news diet of the readers. This would change our view on the ability of the governments to exercise media capture online, implying that they do not need to control all news producers in the market (Besley and Prat, 2006) and instead need to invest in the quality of a handful of controlled outlets. Second, the distribution of consumer tastes for ideological bias in the news is interesting in itself; it provides a revealed-preference measure of the ideological views of the news consumers in a country where surveys might be unreliable. Moreover, the mechanism behind preferences for the ideological bias allows us to understand the fundamental principles driving news consumption and the nature of competition in the online news markets (Xiang and Sarvary, 2007).

At the same time, separating out consumer preferences for the ideological bias in the news from the persistent preferences for the news outlets is challenging. A few existing identification strategies rely on the variation that is not available in the online markets.[4] To overcome this challenge, we propose a new identification strategy that exploits exogenous shifts in the amount of government-sensitive events, ones reflecting ideological positions of the outlets, that happen over time. Intuitively, on days with no sensitive events to report, both the GC and independent news outlets would cover only non-sensitive news, so their news product would not contain any ideological bias. This implies that on such days, the ideological preferences of consumers would not matter for their outlet choice, and they choose

https://www.economist.com/blogs/graphicdetail/2016/02/daily-chart-4.

[4]Gentzkow and Shapiro (2010) use spatial variation in political preferences of the newspapers' consumers; Martin and Yurukoglu (2017) use the variation in channel positions across local cable systems together with the variation in the ideological position of the MSNBC channel over time.

the news outlet only based on their persistent preferences for the outlets. 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. Consumers who prefer the pro-government bias in the news would be more likely to navigate to the GC news outlets. "Conscientious" consumers would be more likely to read ideologically-diverse news outlets, as they value knowing multiple opinions about the events. Consumers who read slanted news only for entertainment would be more likely to go to the extremely ideologically-biased outlets, regardless of the valence of slant on these outlets. In Section 2, we capture this intuition in a stylized demand model.

We focus our analysis on the online news market in Russia. We start with characterizing the level of government control of the top 48 online news outlets that reported in Russian in April 2013 - April 2015. Using information on the ownership structure (Djankov et al., 2003) and reports of alleged government influence, we label the outlets as GC (direct ownership of the government), independent (international or journalists' ownership with no reports of alleged government influence), potentially influenced (owner is an oligarch linked to the government or there are reports of government influence), international and Ukrainian (located outside of Russia and specifically in Ukraine). For these news outlets, we collect all (3.9 million) publication records during this time period.

We use the outlet classification and publication records to find and characterize sensitive news topics with pro-government bias. For this, we compare the news coverage of the GC and independent outlets, looking for systematic differences in the news topics reported and language used. To find these differences, we design a novel and simple classification method that searches for words and phrases that are under- or overused by all the GC news outlets compared to all the independent outlets.

We apply this method to our classification task and find two distinct government-sensitive news topics. First, we find a set of news events that are censored by the GC news outlets. These events mainly correspond to corruption, protests and other political opposition, all of which are internal issues for the country. We thus label these news as "internally sensitive." Interestingly, we do not find substantial difference in the language used by the GC and independent outlets when covering the internally sensitive news, indicating that censorship is the main method of government control of these news topics. Second, we find significant language differences in the GC outlets' coverage of the Ukraine-crisis events of 2013-2015. For this news topic, we compare the coverage of the GC and Ukrainian news outlets to find both pro-Russia and pro-Ukraine ideological slant. The pro-Russia slant of the GC outlets is characterized by the negative framing of Ukraine and positive framing of the pro-Russian separatists. For example, the GC outlets report that Crimea "reunited" with Russia and that the new Ukrainian government is "fascist," "anti-Russian," and conducts a "punitive" military operation in eastern Ukraine. The pro-Ukraine slant of the Ukrainian outlets is quite the opposite; the Ukrainian outlets report that Russia is an "aggressor" country that has "annexed" Crimea and that the Ukrainian government conducts an "anti-terrorist" operation in eastern Ukraine. These language differences fit well with the reports of independent journalists monitoring the news coverage of the Ukraine crisis, validating our results. We use pro-Russia and pro-Ukraine slant to construct the measures of valence (difference between pro-Russia and pro-Ukraine slant usage) and volume (sum of the two) of slant in the news reporting.

The sensitive news classification provides two important ingredients for our identification 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 slant positions of the news outlets. We show that news outlets hold relatively stable reporting and slant positions, always covering a certain percentage of sensitive news and having a certain percentage of articles about the Ukraine crisis with pro-Russia and pro-Ukraine slant. This stability shows a limited supply-side reaction to changes in the relative importance of sensitive news and validates our identification strategy. We approximate the ideological positions of the news outlet by their average share of reporting about sensitive news and average share of the news articles with pro-Russia and pro-Ukraine slant in the Ukraine-crisis coverage.

The final ingredient of our empirical strategy is the news consumption data. We construct a long panel of news consumption occasions in the online news market in Russia using the browsing records from Internet Explorer (IE) Toolbar data for the period of November 2013 - April 2015. Given that the IE Toolbar users might be not fully representative of an average internet user in Russia, in Section 3.3.1 we compare their browsing behavior to the population average. The data suggests that the IE users are older and less interested in entertainment websites, and have higher visit shares of business-focused news websites. At the same time, the IE Toolbar data closely tracks the population in the main identifying variation of our empirical strategy, changes in the news outlets' consumption over time (average correlation of 85.8%), suggesting that the ideological preferences of consumers that we find should match the population's preferences. We also point out that the IE Toolbar users' demographics and browsing records suggest a higher preference of these users for the pro-government ideology compared to an average internet user in Russia, which should bias our results in the opposite direction from what we find.

Using the results above, we estimate the news demand model on a sample of 52,568 frequent news consumers in our data. The preference estimates reveal that an average consumer in the Russian online news market has a distaste for the pro-government ideological positions of the GC outlets but a high persistent preference for these outlets. There is substantial heterogeneity in consumer preferences, with 37.2% of consumers preferring the ideological slant of the GC news outlets in coverage of the Ukraine crisis and 42.11% of consumers preferring lower coverage of the GC news outlets about the internally sensitive news. However, consumers' demand for the GC news outlets is driven primarily by their persistent preferences for these outlets. If the GC news outlets had the ideological positions of the independent outlets, the GC news outlets would get a 20.3% higher market share, corresponding to a rough back-of-the-envelope estimate of an additional \$18.41 million in advertising revenues. This lost advertising revenue pales in comparison to the \$1.21 billion of government subsidies to mass media in Russia in 2015 alone, suggesting that it is relatively inexpensive for the government to compensate the controlled outlets.[5] This difference in the market shares is also compensated by high persistent preferences of consumers for the GC outlets; if the average persistent preferences of consumers for the GC outlets was the same as for the independent outlets, the GC outlets would get a 44.6% lower market share. Such importance of the persistent preferences suggest that the government can effectively manipulate the ideological news diet of the loyal consumers of the controlled outlets, at least in the short run.

Finally, structural demand estimates allow us to examine the mechanism behind consumer preferences for the ideological slant. Only a minority of consumers in the sample, 25.5%, prefer more ideologically-diverse news sources on days with more Ukraine-crisis news, suggesting that they are "conscientious" consumers. An average consumer also prefers the Ukraine-crisis news with a lower volume of slant, although there is substantial heterogeneity in consumer preferences. Both of these results support the theory that consumers prefer slanted sensitive news primarily because of the preference for like-minded news.

[5] Source: `http://www.rbc.ru/politics/29/06/2015/55912ffa9a7947453982cda9`. We use the exchange rate of the end of 2014, which was 60 rubles per dollar. The total of 72.6 billions rubles includes subsidies to the television and print media.

We believe that the primary contribution of this paper is the new identification strategy for consumers' ideological preferences in the news coverage, adding to the empirical literature on news consumption under the ideological slant (Gentzkow and Shapiro, 2010; Martin and Yurukoglu, 2017). Our model of the online news demand builds on 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., 2017). To our knowledge, we are first to estimate a demand model for news allowing for flexible heterogeneity in consumer preferences. Our demand estimates 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). Finally, we propose a new and simple method to measure media slant (Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010; Gentzkow et al., 2016), and to our knowledge we are the first to separate out media bias into censorship, valence and volume of slant, building on the consumer reviews literature (Chevalier and Mayzlin, 2006; Liu, 2006).[6]

The next section builds a stylized model of the news supply and demand and lays out our identification strategy. Section 3 describes the Russian online news market, our data sources and classification of the government-sensitive news. In Section 4 we characterize the reporting of the news outlets and present some suggestive evidence on the direction of consumer preferences. We describe our empirical specification in Section 5 and present the demand estimates and counterfactual simulations in Section 6. Section 7 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.

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

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 the 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. For now, we 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

$$U_{it0} = \epsilon_{it0},$$

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

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), consumer demand for news outlets' products $\{D_A, D_B\}$ 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 when there are more events related to this topic. 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 the amount of the events of the same type that happen on day t, $\{V_t^S, V_t^{NS}\} \in [0, 1]$. We further assume that $c_t^S(x_{jt}^S, V_t^S)$ and $c_t^{NS}(x_{jt}^{NS}, V_t^{NS})$ are convex in x_{jt}, capturing the intuition that it is increasingly costly to discover extra news events of a certain type.[7]

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

[7] Note that we have assumed away fixed costs and common news production costs. Generalizing the model to add these costs does not change the results of the analysis below.

imposes additional costs of production of sensitive news on the outlet A, $c^G(x^S_{At})$, exercising censorship.[8]

Facing this demand and costs structure, the news outlets engage in a simultaneous game and choose the optimal production levels of the commodities x^S_{tj} and x^{NS}_{tj} to maximize their profit functions,

$$\text{Profit}_{At}(x^S_{At}, x^{NS}_{At}) = pD_A(x^S_{At}, x^{NS}_{At}; x^S_{Bt}, x^{NS}_{Bt}) - c^S_t(x^S_{At}) - c^{NS}_t(x^{NS}_{At}) - c^G(x^S_{At}),$$

$$\text{Profit}_{Bt}(x^S_{Bt}, x^{NS}_{Bt}) = pD_B(x^S_{Bt}, x^{NS}_{Bt}; x^S_{At}, x^{NS}_{At}) - c^S_t(x^S_{Bt}) - c^{NS}_t(x^{NS}_{Bt}),$$

where p is the monetary benefit that the company gets from supplying one unit of the product.[9] This game can be solved for $\{x^{S*}_{jt}, x^{NS*}_{jt}\}$ for particular demand and costs specifications following the standard product differentiation literature (Hotelling, 1929; Tirole, 1988).

Two observations follow from this setting. First, the controlled outlet A produces less sensitive news than the independent outlet B, $x^{S*}_{At} \leq x^{S*}_{Bt}$, as it faces higher marginal costs of sensitive news production.[10] Second, the difference in the amount of sensitive news produced by outlets A and B, $x^{S*}_{Bt} - x^{S*}_{At}$, depends on the amount of sensitive events that happen on day t, V^S_t. Unless the government mainly cares about the first few sensitive stories reported by the outlet A (concave c^G), we would expect $x^{S*}_{Bt} - x^{S*}_{At}$ to be increasing in V^S_t. Intuitively, when there are no sensitive news to report, $V^S_t = 0$, it is very costly for both news outlets to produce sensitive news (high c^S_t), so both outlets produce very low x^{S*}_{jt} and an extra cost of c^G for an outlet A does not matter that much. In contrast, when there are a lot of sensitive news to report (high V^S_t), the cost of sensitive news production is very low (c^S_t close to zero), so c^G plays a more important role.[11]

Our key identification argument relies on the second observation above. In Section 4.1, we show that the difference in the sensitive news reporting between the GC and independent

[8]For 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).

[9]For simplicity, we assume that news outlets do not have control over p. For example, p can be the cost-per-impression (CPM) rate the outlets get from displaying ads on their webpages. The CPM rates are often determined by competition among content producers in the advertising market.

[10]Given that the government determines c^G, it effectively chooses x^{S*}_{At} based on its objective function.

[11]More formally, this relationship comes from the assumptions that $c^S_t(x^S_{jt}, V^S_t)$ is decreasing in V^S_t and convex in x^S_{jt}. On days with higher V^S_t, the optimal level of x^S_{jt} production is higher for both outlets. When the x^{S*}_{jt} is higher, both outlets are on the less steep part of the profits curve (which is concave as c^S_t is convex), so in order to equate extra marginal costs coming from the government control, c^G, the outlet A should give up a larger amount of sensitive news reporting, x^S_{At}, than the outlet B.

outlets indeed increases with V_t^S. We use these changes in the sensitive news reporting induced by V_t^S to identify consumer preferences for sensitive news, β.

2.2 Extensions

We now extend and adjust the basic model to account for other important features of the online news consumption and production.

Persistent preferences. Apart from the news commodities supplied, outlets can differentiate themselves in a variety of ways, such as website design, overall quality of the news coverage, etc. Consumers can like or dislike these attributes of the outlets,

$$U_{ij} = \alpha_{ij} + \beta_i x_j^S + \lambda_i x_j^{NS} + \epsilon_{ij} \;:\; j \in \{A, B\}, \{x_j^S, x_j^{NS}\} \in [0,1],$$

where α_{ij} represent the matching value between consumer i's preferences and features of the news outlet j.

Space constraints. Up to this point we have assumed that news outlets make two separate choices of x_{tj}^S and x_{tj}^{NS} that only depend on the realizations of V_t^S and V_t^{NS}. In practice, outlets operate under the 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, $x_{tj}^S + x_{tj}^{NS} = 1$, so the only thing that varies over time is the ratio of the produced sensitive and non-sensitive news commodities.[12]

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

$$U_{ij} = \alpha_{ij} + \beta_i x_j^S + \lambda_i(1 - x_j^S) + \epsilon_{ij} = (\alpha_{ij} + \lambda_i) + (\beta_i - \lambda_i)x_j^S + \epsilon_{ij},$$

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.[13] With a slight abuse of notation, we redefine consumer utility to get rid of γ_i,

$$U_{ij} = \alpha_{ij} + \beta_i x_j^S + \epsilon_{ij},$$

[12] We use this assumption in the empirical part since we observe only the relative importance of sensitive and non-sensitive news, V_t^S and V_t^{NS}, proxied by the share of sensitive and non-sensitive articles in the overall news market.

[13] Note 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.

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 ideological stand bias, such as supporting, opposing, or being neutral about the government.[14]

To account for this, we extend the model by allowing the news outlets to choose the valence and volume of the ideological slant in their sensitive news reporting. The valence of slant, val_{jt}, reflects the ideological framing of sensitive news in the outlet's reporting; for example, referring to the Russian government is an "aggressor" frames Russia negatively in the Ukraine-crisis coverage (anti-Russian-government slant), while calling the Ukraine government "fascist" frames Ukraine negatively (pro-Russian-government slant). The volume of slant, vol_{jt}, captures the idea that news outlets can decide to abstain from any slanted language, or can use both pro- and anti-government slant in their reporting to make the news more emotional and "entertaining" (Thussu, 2008).[15]

Consumers hold stable preferences for the valence and volume of slant in the sensitive news reporting,

$$U_{ij} = \alpha_{ij} + (\beta_i + \gamma_i^{\text{val}} \text{val}_{jt} + \gamma_i^{\text{vol}} \text{vol}_{jt}) x_j^S + \epsilon_{ij}.$$

The preference for the valence of slant, γ_i^{val}, captures consumer's preference for the ideology of the reporting, driven either by a preference for the like-minded news (Klayman, 1995) or an interest in a variety of news opinions (Mullainathan and Shleifer, 2005). The preference for the volume of slant, γ_i^{vol}, captures consumer's preference for more or less sensational news (Thussu, 2008).

While it is natural to assume that framing of sensitive news is costless for the news outlets (Gentzkow and Shapiro, 2010), we stop short of incorporating the framing decision in the outlets' news production. Instead, we resort to the empirical analysis in Sections 3.2 and 4.1, where we describe the nature of the ideological slant in the sensitive news reporting and show that the share of slanted news articles about sensitive topics is stable over time.

[14] The literature refers to this ideological bias as 'framing and ideological stand bias" (Prat and Strömberg, 2013) and "distortion of news" (Gentzkow et al., 2016).

[15] Standard measure of slant include only one dimension of valence (Gentzkow and Shapiro, 2010). We bring the idea of the volume of slant from the consumer reviews literature (Chevalier and Mayzlin, 2006; Liu, 2006), where it applies to the product ratings.

Conscientious news consumption. The final goal of the model is to separate out consumers who prefer the ideologically-slanted sensitive news because of the confirmation bias (Klayman, 1995) from the "conscientious" news consumers (Mullainathan and Shleifer, 2005). To test whether consumers are conscientious or prefer like-minded news, we exploit the occasions of multiple news outlet consumptions within the same day. If consumers prefer like-minded news, the outlets they read on days with a lot of sensitive news should be "concentrated" in terms of their ideology, meaning that they should have a similar valence of slant of the sensitive news reporting. In contrast, conscientious consumers are interested in a variety of opinions, so their choice set on the days with a lot of sensitive news should be less concentrated compared to a day with few sensitive news reported.

Borrowing from the literature on variety-seeking behavior in the product choice (McAlister and Pessemier, 1982; Kim et al., 2002), we can incorporate this idea into the consumer utility:

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

where τ is the choice occasion of consumer i on day t and $s_{i\tau}^{\text{val}}$ is the valence of slant of the outlet that consumer reads on $\tau - 1$. Consumer preference for the ideological variety-seeking, ρ_i, captures whether the consumer starts reading more ideologically-concentrated news as x_{jt}^S increases (negative ρ_i), suggesting that he prefers the like-minded news, or less concentrated news (positive ρ_i), behaving like a conscientious type.[16]

2.3 Identification

Our identification strategy of consumer preferences relies on exogenous shifts in the amount of sensitive news that happens over time. The key observation is that there will be more reporting about sensitive news when more sensitive events happen, making the ideological positions of the news outlets more important in consumers' outlet choice problem. In Section 4.1, we validate this empirical strategy by showing that the ideological positions of the outlets are stable over time, meaning that the outlets tend to report a certain share of sensitive events that happen every day and have a certain share of slanted articles in their

[16]We 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 and Sarvary, 2007) models.

sensitive news reporting.

Shifts in the importance of ideological positions of the news outlets identify consumer demand parameters. There are five coefficients of interest in the model. The distributions of the persistent preferences of consumers, α_j, are identified from their outlet choices when there is no sensitive news reported, $x^S_{tj} = 0$. The distributions of the relative preference for the sensitive news, β, valence of slant, γ^{val}, and volume of slant, γ^{vol}, are identified from shifts in the sensitive news reporting, x^S_{tj}, induced by changes in V^S_t, and slant positions of the news outlets, val_j and vol_j. Finally, the distribution of ρ, an ideological variety-seeking preference of consumers, is identified from changes in x^S_{tj} and the distance between the slant valence of two subsequently chosen news outlets within a day.

3 Data

In this section we describe the state of the Russian online news market, publication records and browsing datasets used, and our classification of the government-sensitive news.

3.1 Online News Market in Russia

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. Since the beginning of 2013, political pressure has forced a number of top online news outlets to remove their chief editors.[17] The most prominent examples include the dissolution of RIA Novosti, a state news agency known for balanced news coverage under its editor-in-chief Svetlana Mironyuk, in December 2013[18] and the layoff of Galina Timchenko, editor-in-chief of one of the top online news outlets in Russia, lenta.ru, in March 2014.[19] Government control intensified in February of 2014 with the Ukrainian crisis and the annexation of Crimea, with the government blocking websites of some opposition leaders in March 2014[20] and implementing a law to limit the foreign

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

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

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

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

ownership of Russian news outlets to 20% as of January 2016.[21]

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

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

Table presents the simplified domain names; for example, 1tv stands for www.1tv.ru. Most domains have the www.*.ru structure, with some exceptions. The classification is done based on the media ownership information, evidence of the indirect influence such as removing news articles due to political pressure, and interviews with media professionals. Online Appendix 9.1 presents more detailed information on the ownership structure and evidence of the indirect influence for each news outlet.

At the end of 2014, the online news media landscape in Russia still included both groups of GC and independent news outlets. However, an increasing number of the news outlets that are indirectly influenced forces us to create a separate classification group, which we call "potentially influenced." This group includes the news outlets that are not owned or directly influenced by the government but that can face some political pressure indirectly, e.g. through their owners. In addition, we create two separate groups of the prominent international and Ukrainian news outlets with Russian language news coverage.

Table 1 presents the top 48 Russian-language news outlets by groups.[22] The classification is done based on the media ownership information, evidence of the indirect influence such as removing news articles because of the political pressure, and interviews with media professionals.[23] The first column contains the GC news outlets, ones that are directly

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

[22] We 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 dw.com.

[23] Online Appendix 9.1 presents more detailed information on the ownership structure and evidence of the

owned by the government or members of the incumbent political party. The second and third columns include the "potentially influenced" news outlets, ones that could be influenced by Kremlin indirectly. This group includes any news outlets that are "suspicious," e.g. owned by an oligarch close to the Kremlin or reported to have removed news articles after a request from the Kremlin. Given the ambiguous nature of the government control of these news outlets, we do not use them in the sensitive news classification stage. The fourth column 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. Finally, columns five and six present international and Ukrainian news outlets with Russian language news coverage.

3.2 Publication Records Data

For the 48 outlets described above, we collect information on publications for the period starting April 1, 2013, and ending March 31, 2015. The data are collected directly from archives on news outlet websites and from the media archives *medialogia.ru* and *public.ru*. The resulting panel contains 3.9 million publications. For each article, we collect the title, text,[24] URL link, and timestamp. Table 2 presents the number of articles per type of news outlet. Online Appendix 9.2 provides more information on the publication records data collection and processing.

Table 2: Number of articles by type of news outlet

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

indirect influence for each news outlet, as well as the reasons to be included in each group.

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

3.2.1 Government-Sensitive News

What are the government-sensitive news topics and the corresponding slant? To find such news, we use our knowledge of the ownership structure of the news outlets and their publication records.

Following the topic modeling literature (Blei et al., 2003), we treat the news articles as collections of words or n-grams, which we consider indicative of the certain news topics and slant.[25] Our key observation is that censorship of news topics and language slant should affect all GC news outlets, decreasing the reporting of the corresponding topics and usage of the corresponding language. To capture this idea, we propose a simple classification algorithm:

1. Compute share of usage of a word or n-gram v by a news outlet j : $\text{sh}_{vj} = \frac{\text{count}_{vj}}{\sum_v \text{count}_{vj}}\ \forall\, v, j$, where count_{vj} is a number of occurrences of v in j's coverage;

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

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

4. For each v, compute the difference in ranks between the GC and independent news outlets, $\Delta\text{Rank}_v^{\text{GC-Ind}} = \text{Rank}_v^{\text{GC}} - \text{Rank}_v^{\text{Ind}}$;

5. Repeat steps 1-4 K times with randomly re-assigned counts of word or n-gram usage within each news outlets, leading to K samples of $\Delta_k^{\text{Random}}\text{Rank}_v^{\text{GC-Ind}}$;[26]

6. v is significantly underused by the GC news outlets if its $\Delta\text{Rank}_v^{\text{GC-Ind}}$ is lower the 5% quantile of $min_v \Delta_k^{\text{Random}}\text{Rank}_v^{\text{GC-Ind}}$ across K draws, meaning that this rank difference

[25] We simplify the articles using standard processing techniques such as stemming and removal of stop words. Online Appendix 9.2 provides more information on this processing.

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

occurs by chance very rarely.[27]

First, we apply this algorithm to the news corpus to find sensitive news that are censored by the GC outlets.[28] Following Franceschelli (2011), we approximate the news topics in our corpus by the universe of the n-grams of the proper nouns, words which contain information about actors in the news, toponyms reflecting where the news has happened, etc.[29] We consider all words starting with a capital letter as proper nouns except for the first words in the sentences. Given that censorship is the omission of facts, the proper nouns corresponding to the censored topic will be underused in the GC news outlets' publications.

To find censored news topics, we apply the classification procedure to 21,709 unigrams and 13,514 bigrams of the proper nouns that appear more than 200 times in the news publications in our sample period.[30] Subfigure (a) of Figure 1 presents the resulting distributions of bigrams $\Delta\text{Rank}_v^{\text{GC-Ind}}$ and $\Delta_k^{\text{Random}}\text{Rank}_v^{\text{GC-Ind}}$ for one k.[31] The red line corresponds to the 5% quantile cut-off level, defined in the step 6 of the classification algorithm, which is -18.8. After we exclude the proper nouns related to the profession of journalism and reflecting news production process (such as the names of journalists, news outlets, media owners, etc.), there are 34 bigrams with rank score differences $\Delta\text{Rank}_v^{\text{GC-Ind}}$ below this threshold.[32] To provide an example of the nature of these bigrams, Table 3 presents 10 of them with the lowest rank score difference $\Delta\text{Rank}_v^{\text{GC-Ind}}$. All of these bigrams are the names of actors related to the potentially sensitive issues for the Russian government, supporting the fact that they represent the censored news topics.[33] Given this, we consider any news article

[27]There are multiple benefits of this simple procedure. Using the shares of words instead of counts allows us to normalize sizes of the news outlets. Converting the share of usage to an ordinal rank limits the effect of the outliers, e.g. news outlets that under- or overuse some particular words or n-grams. Classifying the news topics in terms of words and n-grams instead of using dimension reductions (Blei et al., 2003; Cagé et al., 2017) decreases the noise from word co-occurrence and thus increases the probability that differences of the word usage are detected.

[28]We note that our measure of censorship relies on the difference in coverage between the GC and independent news outlets and thus does not account for a potential self-censorship. Our censorship measure is close "state censorship" in the classification of Crabtree et al. (2015).

[29]For 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 the 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").

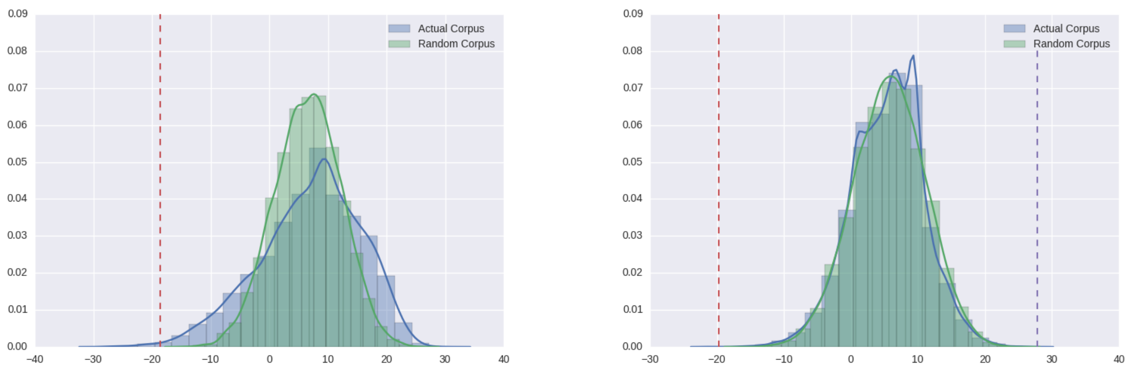
[30]The threshold of 200 times ensures that the proper nouns in the analysis refer to the substantial topics. It is chosen arbitrarily.

[31]We set $K = 500$. Online Appendix 9.3 presents the unigram results.

[32]Overall, there are 54 bigrams with rank score differences $\Delta\text{Rank}_v^{\text{GC-Ind}}$ below the -18.8 threshold. We exclude the names of journalists and news outlet names since these words reflect the news production process (citing sources, excluding links and excerpts, etc.) and not reflecting the censored news. Tables 12 and 13 in Online Appendix 9.3 present all 54 censored bigrams.

[33]There are three prominent opposition politicians, with Alexei Navalny being mentioned twice in two dif-

Figure 1: Histograms of $\Delta\text{Rank}_v^{Ind-Gov}$ across (a) bigrams of the proper nouns in all publications, and (b) non-proper nouns in publications about the censored (internally sensitive) news topics.



(a) Censored News Bigrams (b) Slant in Censored News

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

that mentions one of these 34 bigrams of proper nouns or one of the 10 underused unigrams of the proper nouns[34] as an article about a government-sensitive news topic. We label this group of sensitive news topics as "internally sensitive" because most of the censored proper nouns correspond to the internal issues such as political opposition, protests and corruption.

Is there slant in the internally sensitive news? To check for the language differences in the publications about this sensitive topic, we apply the classification algorithm to non-proper nouns in the news articles about internally sensitive news. Subfigure (b) of Figure 1 presents the resulting distributions of $\Delta\text{Rank}_v^{\text{GC-Ind}}$ and $\Delta_k^{\text{Random}}\text{Rank}_v^{\text{GC-Ind}}$. We find little evidence of slant in the internally sensitive news coverage. Out of 37,734 words in the corpus, only four words are systematically omitted by the GC news outlets, and only one word out of them is

ferent spellings, two close affiliates of Vladimir Putin who are frequently mentioned in the events of potential corruption, Pussy Riot, a band that became famous for its protest activities, Sergei Guriev, a prominent Russian economist who had to flee Russia after a politically-motivated interrogation by government investigators, Svetlana Davydova, a mother-of-seven arrested for a phone call to the Ukrainian embassy that was allegedly an act of treason, and Marat Gelman, a former director of the Perm Museum of Contemporary Art allegedly fired for refusing to remove a controversial political exposition "Welcome! Sochi 2014", `http://www.wiki.ncac.org/Welcome!_Sochi_2014`.

[34]We use unigrams to make sure that we do not exclude facts described with a single proper noun. See Online Appendix 9.3 for more details.

Table 3: List of the top 10 bigrams of the proper nouns underused by GC news outlets.

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

indicative of slant (the word "prisoner" related to the arrested opposition activists).[35] We conclude that we do not find language difference, or slant, in the internally sensitive news reporting by the GC and independent news outlets.

To find a sensitive news topic with ideological slant, we use the news about 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 news outlets controlled by the Russian government.[36] To make sure that we capture all the news about this topic, we consider any news article that mentions the proper noun "Ukraine" to be related to the Ukraine crisis. Interestingly, with the beginning of the Ukraine crisis, the GC news outlets has increased their reporting about Ukraine disproportionally more than the independent news outlets, showing that the overall news about the Ukraine-crisis events was not censored.[37]

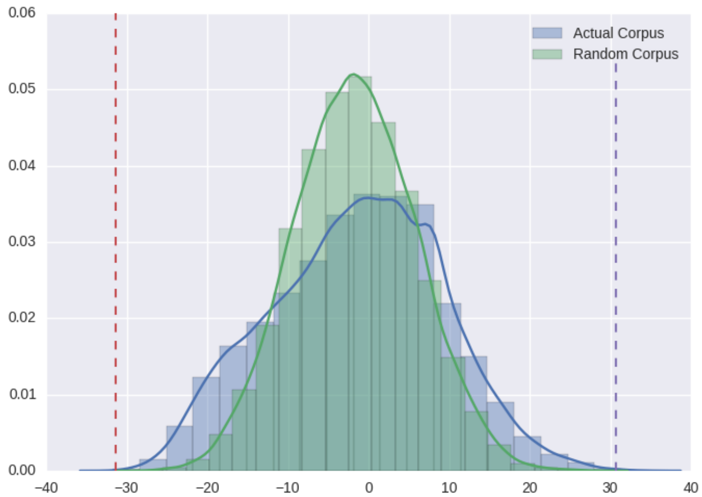
We check the language difference in the Ukraine-crisis coverage by applying the classification procedure to the non-proper nouns in the news articles that mention Ukraine. For this classification, we compare the coverage of the GC and Ukrainian news outlets, since these two types of news outlets should have the most dissimilar views about the topic. Figure 2

[35] Among the other three unigrams are the words "interview" and "editor," related to the means of information delivery, and the word "fired," related to firing of one of the journalists of an independent news outlet. We exclude these words as they are related to the journalism profession and not to the news covered.

[36] For a broad overview, see `https://en.wikipedia.org/wiki/Ukrainian_crisis` and `https://en.wikipedia.org/wiki/Media_portrayal_of_the_Ukrainian_crisis#Media_in_Russia`.

[37] Figure 10 in Online Appendix 9.4 illustrates this by presenting the share of news articles that contain the word "Ukraine" that were published in the independent, government-influenced, and GC outlets over time. The figure also shows that there were only 2-3% of articles mentioning Ukraine before the beginning of the crisis and 20-30% after the beginning of the crisis, validating our definition of the Ukraine-crisis news.

Figure 2: Histograms of $\Delta\text{Rank}_v^{Ukr-GC}$ across words in the Ukraine-crisis news topic corpus.



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

presents the resulting distributions of $\Delta\text{Rank}_v^{\text{Ukr-GC}}$ and $\Delta_k^{\text{Random}}\text{Rank}_v^{\text{Ukr-GC}}$. We find only 13 words that are significantly overused by the GC news outlets and two words that are overused by the Ukrainian outlets. However, the visual differences in the distributions of $\Delta\text{Rank}_v^{\text{Ukr-GC}}$ and $\Delta_k^{\text{Random}}\text{Rank}_v^{\text{Ukr-GC}}$ imply that there is some difference in the language used, suggesting that our classification procedure might be too restrictive in this case.[38]

To find the ideological slant, we combine our classification results with the reports of journalists and fact-checking websites that have described pro-Russian and pro-Ukraine slant in the news.[39] Anecdotally, the pro-Russian slant frames the new Ukrainian government as a "fascist junta" that is conducting a "punitive operation" against the "rebels" in eastern Ukraine, and the pro-Ukraine slant frames Russia as an "aggressor" that has "occupied" the Ukrainian territory and supports "terrorists" and "separatists" in eastern Ukraine. We screen the top 50 words with the lowest and highest rank score differences in our classification for the terms that reflect these ideologically-slanted terms. The language is remarkably consistent. The words with the highest rank difference $\Delta\text{Rank}_v^{\text{Ukr-GC}}$ include the "reunion"

[38]There might be additional noise in the classification since there are only three Ukrainian news outlets in the sample.

[39]We use the resources from `stopfake.org`, a fact-checking website supported by faculty and alumni of the Mohyla School of Journalism and students from the Digital Future of Journalism program in Kyiv, Ukraine.

of Russia with Crimea (ranked 1 highest out of 38,584 words), the "radical" (ranked 2) "anti-Russian" (ranked 6) protesters who have "overturned" (ranked 4) the former government in a "coup" (ranked 7), the "punitive" (ranked 3) operation and "bombing" (ranked 6) of eastern Ukraine, etc. The words with the lowest rank difference $\Delta\text{Rank}_v^{\text{Ukr-GC}}$ include the "annexation" (ranked 2 lowest out of 38,584) and "occupation" (ranked 9) of Crimea by Russia, and the Ukraine army is conducting an "anti-terrorist" (ranked 4) operation against the "separatists" (ranked 24).[40] We pick the words that match the reports of the journalists from the 50 words with the lowest and highest rank score differences and label them as indicative of the pro-Russia and pro-Ukraine slant. The resulting set contains 15 words that correspond to the pro-Russia slant and seven words that correspond to the pro-Ukraine slant in the Ukraine crisis.[41] Table 16 in Online Appendix 9.4 contains the final list of the selected words. We denote the articles that contain both the word "Ukraine" and one of the selected pro-Russia- or pro-Ukraine-slanted words as an article about the Ukraine crisis with the pro-Russia or pro-Ukraine slant, respectively.

3.3 News Consumption Data

To measure news consumption, we use the Internet Explorer (IE) Toolbar browsing data, which includes complete browsing histories for a subset of IE users. The users included in the IE Toolbar data have installed a plug-in on their IE and opted-in for the data collection.[42] 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.[43]

Although IE Toolbar data are collected for several years, the unique user IDs are kept only for one and a half years. By the time the data collection was conducted, the earliest available browsing data with user IDs were from November 15, 2013. We thus collect the browsing data between November 15, 2013, and March 31, 2015,[44] for all users with the IE

[40]Table 15 in Online Appendix 9.4 presents the top 10 overused words by the GC and Ukrainian news outlets in the Ukraine-crisis news coverage.

[41]We validate our classification with four independent research assistants, who go through the list of the potentially slanted words and label them as ideologically slanted.

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

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

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

language set to Russian.

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

To understand the online news consumption in the IE Toolbar data, we classify the webpages that consumers visit into four groups: main pages of the websites, news subdirectories, news articles, and other pages (such as special projects, photos, videos, etc.).[45] Table 4 shows summary statistics of browsing by types of webpages. News articles account for 39.3% of the page views on news websites. News directories and subdirectories account for another 36%, with other pages accounting for 24.7%.

Table 4: Summary of browsing behavior

	Page views	% of Page Views
Main page	5,344,041	20.1%
News articles	10,420,780	39.3%
News subdirectories	4,225,221	15.9%
Other	6,584,713	24.7%
Total	26,537,267	100%

3.3.1 IE Toolbar Representativeness

Before we proceed with the analysis, we test whether the news consumers in the IE Toolbar data are representative of the overall population of news consumers in Russia. To make this comparison, we collect data on the number of daily visits for a subset of news outlets in Russia using `liveinternet.ru` (LI), a website that tracks statistics for the Russian internet. We use the digital archive Wayback Machine to collect historical data on website usage. Due to the layout of the website ranking on LI, we have reliable records of usage over the period of time studied for the top 30 websites on the Russian internet, which includes seven news

user IDs.

[45] Online Appendix 9.5 contains details on classification.

websites from our sample.[46]

Table 17 in Online Appendix 9.6 compares browsing habits of the news consumers in the IE Toolbar data to the general population. IE Toolbar consumers tend to be older,[47] more interested in the weather and less interested in streaming and entertainment websites. However, the shares and rankings of the website are relatively similar, suggesting that the IE Toolbar consumers are not too different from the general population in Russia.

Table 5 zooms into the seven news websites that have reliable records in the historical LI data, comparing the visit shares and rankings of the news outlets. The results are mixed. On the one hand, five out of the top seven news outlets in the LI data are also present in the top seven in the IE Toolbar data. On the other hand, there are a couple of significant deviations, with the second outlet in the LI data, ria.ru, ranking 14 in the IE Toolbar data, and the market leader, rbc.ru, having a substantially higher visit share in the IE Toolbar data. One of the potential explanations for these differences is the anecdotal over-representativeness of the office workers in the IE Toolbar data, supported by older demographics of the IE Toolbar. This can explain both a higher visit share of the rbc.ru (it is a news agency with a more extensive business news coverage) and a lower visit share of the ria.ru (it is another news agency competing with rbc.ru).

Table 5: Comparison of website rankings in IE Toolbar and LI.ru

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

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

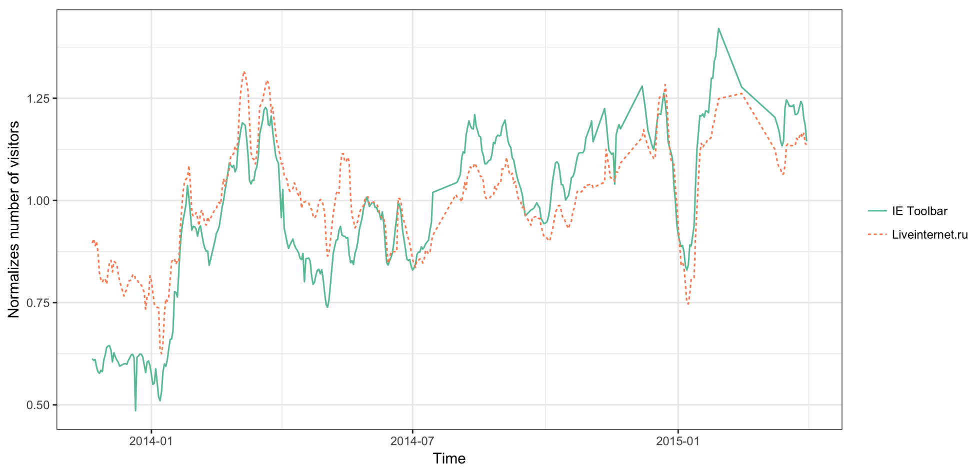
For a better understanding of the news consumption patterns in the IE Toolbar data, we examine changes of the news outlets' traffic in the IE Toolbar and LI datasets, the key variation in our empirical strategy. Figure 3 presents the average traffic to the top seven

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

[47]They are more likely to visit odnoklassniki.ru, a social network with older demographics, and less likely to visit vk.com, a social network with younger demographics than the general population.

LI news outlets based on the LI and IE Toolbar data.[48] 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 11 in Online Appendix 9.6 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. We conclude that while there are some browsing and news consumption differences between the IE Toolbar and the population, consumption habits of the IE Toolbar users are informative about the news consumption of the population.

Figure 3: 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.

4 Descriptive Evidence

In this section we characterize the reporting of the news outlets and present some model-free evidence that suggests the direction of consumer preferences.

4.1 Reporting about the Sensitive News

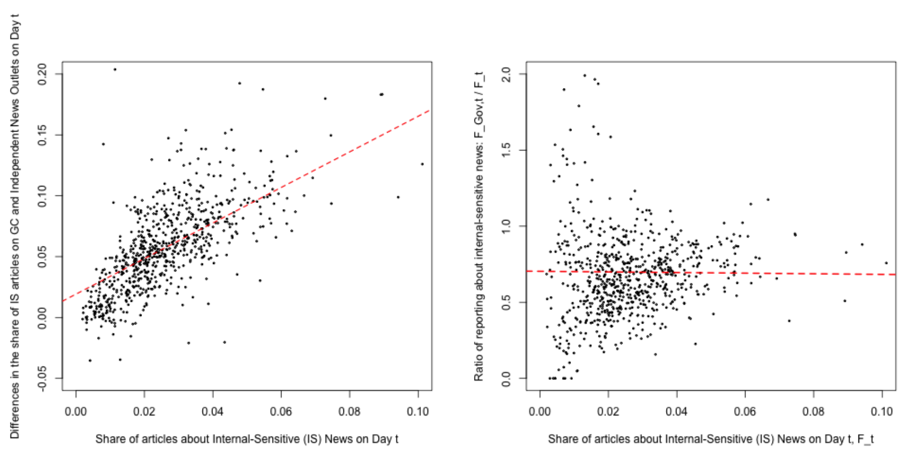
In Section 3.2, we have identified two groups of the news topics that are sensitive for the government: internally sensitive news and Ukraine-crisis news. For our identification strategy, we need a measure of the amount of sensitive news that happens every day. While we

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

do not observe the actual amount of sensitive events that happen daily, we do observe the overall number of articles about the topic l published every day, $\sum_j N^{l}_{tj}$. We use $\sum_j N^{l}_{tj}$ to compute the overall share of articles about the topic l on day t, $F^l_t = \frac{\sum_j N^l_{tj}}{\sum_l \sum_j N^l_{tj}}$, which is a measure of the relative importance of the topic on this day. The relative importance measure is more appropriate if we use the published articles since the news outlets are operating under capacity constraints.[49]

We next examine the reaction of the news outlets to the changes in the relative importance of the sensitive news. First, we focus on the internally sensitive news. We test whether there is a higher difference in the reporting of the independent and GC news outlets during the days with more internally sensitive news, as the basic model in Section 2.1 predicts. Subfigure (a) in Figure 4 shows the relationship between the difference in the daily share of the internally sensitive news articles on the independent and GC news outlets, $F^{IS}_{t,Ind} - F^{IS}_{t,GC}$, and relative importance of the internally sensitive news on this day, F^{IS}_t. There is a strong positive correlation between the two, supporting that the censorship becomes more important on the days with more internally sensitive news events.

Figure 4: GC news outlets censor more news on the days with a lot of internally sensitive news.



(a) Difference $F^{IS}_{t,Ind} - F^{IS}_{t,GC}$ (b) Ratio $F^{IS}_{t,GC}/F^{IS}_t$

The red line corresponds to the fitted values of the linear regression. Subfigure (a) corresponds to the linear regression of $F^{IS}_{t,Ind} - F^{IS}_{t,GC}$ on F^{IS}_t; the slope coefficient is statistically significant ($p < .001$). Subfigure (b) corresponds to the linear regression of $F^{IS}_{t,GC}/F^{IS}_t$ on F^{IS}_t; the slope coefficient is not statistically significant ($p = .822$).

[49]Capacity constraints imply that if the news outlet publishes more news on the topic l, they automatically have to decrease the number of publications about the other news topics, which we do not want in the our news importance measure.

How does the fraction of censored news differ with the importance of this sensitive news topic? Subfigure (b) in Figure 4 shows that the share of internally sensitive news that the GC news outlets report does not differ significantly with the relative importance of sensitive news, F_t^{IS}. On average, the GC news outlets report 30.2% less news than the average number of articles in the news market on that day and 69.7% less news articles than the independent news outlets report. We further check if the news outlets adjust the ratios of the internal sensitive news reporting when there is more sensitive news and find that interactions of outlet fixed effects with changes in F_t^{IS} explain only 1% of variation in the reporting ratios, while fixed effects of the outlets explain around 30% of the variation.[50] Given these limited changes in the share of the reporting of the sensitive news over time, we approximate the ideological positions of the news outlets by their average reporting about the sensitive news, $\bar{F}_j^{IS} = \frac{\sum_t F_{tj}^{IS}}{\sum_t \sum_l F_{tj}^l}$. Figure 12 in Online Appendix 9.7 presents the resulting ideological positions of the news outlets in their reporting of the internally sensitive news.

We find similar stable ideological positions in the Ukraine-crisis news reporting. Using the data since the beginning of the Ukraine crisis, we find that news outlet fixed effects explain 81.7% of the variation in the share of Ukraine-crisis news reported by the news outlets, 27% of the variation in the share of pro-Ukraine slanted news and 33.6% of the variation in the share of pro-Russia slanted news, while the interactions of outlet fixed effects and amount of Ukraine-crisis news explain only 3%, 0.2% and 1%, respectively.[51] We thus approximate the ideological positions of the news outlets by their average reporting about the Ukraine-crisis news, $\bar{F}_j^{Ukr} = \frac{\sum_t F_{tj}^{Ukr}}{\sum_t \sum_l F_{tj}^l}$, and the average share of slanted news articles in the Ukraine-crisis news reporting, $\frac{\sum_t N_{tj}^{\text{pro-Russia}}}{\sum_t N_{tj}^{\text{Ukr}}}$ and $\frac{\sum_t N_{tj}^{\text{pro-Ukraine}}}{\sum_t N_{tj}^{\text{Ukr}}}$.

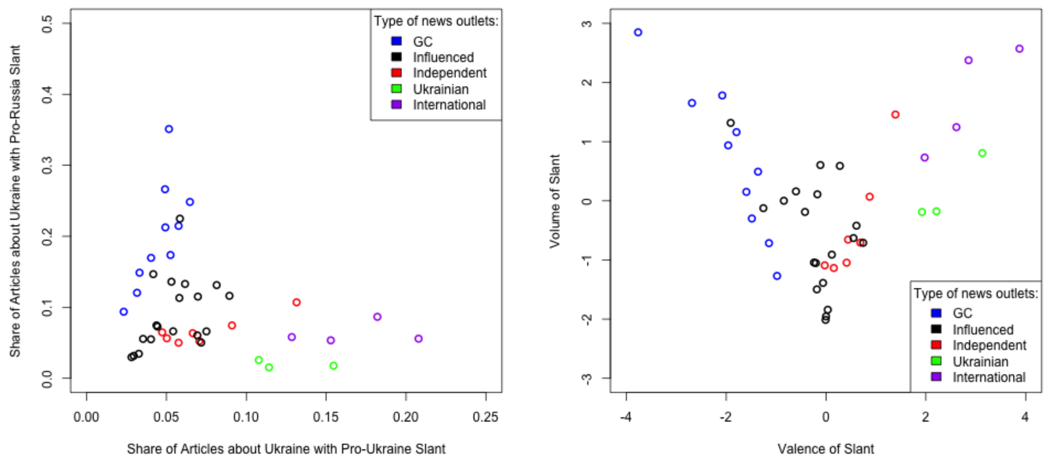
Finally, we use the share of the Ukraine-crisis articles with pro-Russia and pro-Ukraine slant to construct the measures of valence and volume of slant in the Ukraine-crisis news reporting. Subfigure (a) in Figure 5 presents the average shares of articles about the Ukraine crisis that have at least one pro-Russia or pro-Ukraine slant word, $\frac{\sum_t N_{tj}^{\text{pro-Russia}}}{\sum_t N_{tj}^{\text{Ukr}}}$ and $\frac{\sum_t N_{tj}^{\text{pro-Ukraine}}}{\sum_t N_{tj}^{\text{Ukr}}}$. We construct the measure of valence and volume of slant from these nor-

[50]The regression of the ratios $\frac{N_{tj}^{IS}}{\sum_j N_{tj}^{IS}}$ on outlet fixed effects gives an adjusted R-squared of 0.3069, and the same regression with an interaction of outlet fixed effects with F_t^{IS} – an adjusted R-squared of 0.3172.

[51]The first numbers correspond to the adjusted R-squared in the regressions of $\frac{N_{tj}^{IS}}{\sum_j N_{tj}^{IS}}$ on outlet fixed effects or outlets fixed effects and their interactions with the volume of Ukraine-crisis news. The second and third numbers correspond to the adjusted R-squared in the regressions of the share of slanted articles (contain pro-Ukraine or pro-Russia slanted words) in the outlet's daily reporting on outlet fixed effects or outlets fixed effects and their interactions with the volume of Ukraine-crisis news.

malized shares.[52] The volume of slant, vol_j, is defined as the sum of normalized shares, and the valence of slant, val_j, is defined as the difference of the pro-Ukraine slant and pro-Russia slant. Subfigure (b) of the Figure 5 presents the resulting ideological positions of the news outlets. We can see that the majority of the independent and potentially influenced news outlets are neutral, with some independent outlets slightly leaning to pro-Ukraine and some potentially influenced outlets leaning slightly pro-Russia. There are also neutral news outlets with different volume of slant. Figures 13 and 14 in Online Appendix 9.7 present the ideological positions of the news outlets with their name labels.

Figure 5: Volume and valence of slant of the news outlets in the Ukraine-crisis news coverage.



(a) pro-Russia and pro-Ukraine Slant (b) Valence and Volume of Slant

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

4.2 Changes in News Consumption

Before we get to the full empirical specification, we explore some model-free evidence that can suggest the direction of consumer preferences for the ideological position of the government.

Our empirical exercise is fairly simple. Above, we have shown that news outlets have stable ideological positions, which we characterize by the average share of reporting of the sensitive news and average valence and volume of slant. These ideological positions become

[52]We normalize the shares mean to 0 and standard deviations to 1 to make them comparable.

more important for consumers on the days when there are more sensitive news events, as suggested by the stylized model that we present in Section 2. To understand consumer preferences for these ideological positions, we explore how the market shares of the news outlets change with the volume of sensitive news.

We construct the market shares of the news outlets using the news consumption records in the IE Toolbar data. To make sure that we capture only actual news consumption, we define a news consumption of an outlet j on day t by consumer i as navigation to at least one news article on j during t. 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.

We then regress the market share changes of each outlet j on the relative importance of sensitive news and some controls,

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

where F_t^{IS} and F_t^{Ukr} correspond to the share of articles about internally sensitive news and Ukraine-crisis news, and Z_{jt} corresponds to the controls, such as indicator variables for weekdays and time trends.[53] The slope coefficients b_j^{IS} and b_j^{Ukr} correspond to the change in the market shares due to the change in the amount of sensitive news in the market.

We estimate b_j^{IS} and b_j^{Ukr} for 42 news outlets including weekday and week indicator variations as controls.[54] Figure 6 summarizes and visualizes the estimation results. Each point on the subfigures (a)-(c) represents an estimate of b_j^{IS} or b_j^{Ukr} for the news outlet j. Points of larger size represent a larger absolute value of the estimates, with blue and red colors corresponding to positive and negative estimates of b_j^{IS}, respectively. Points with bold borders represent outlets with statistically significant estimates of b_j^{IS}.[55]

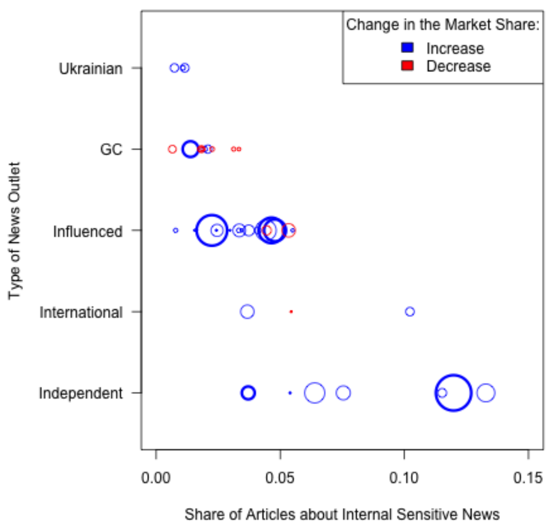
Subfigure (a) of Figure 6 visualizes the estimates of b_j^{IS}. Results suggest that news outlets with more reporting about the internally sensitive news are more likely to get an increase in the market shares on the days with more news about the internally sensitive events. We test this more formally by regressing the b_j^{IS} estimates on $\bar{F}_j^{IS}$, the average share of reporting about the internally sensitive events by the news outlets j. Table 6 presents the

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

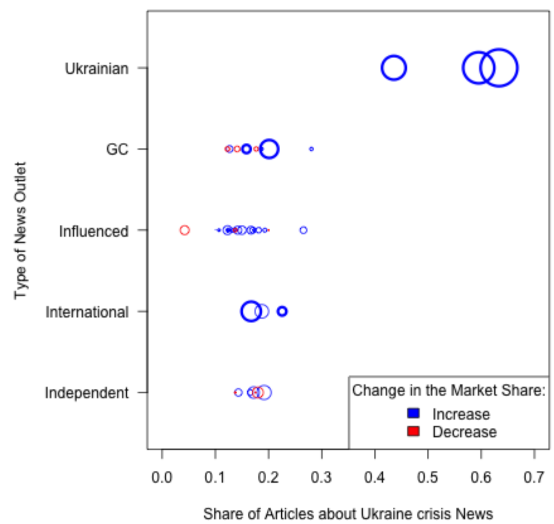
[54]We exclude five news outlets for which we do not have information about the text of the articles, and one news outlet (`dw.de/ru`) for which we have few (10) news consumption occasions.

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

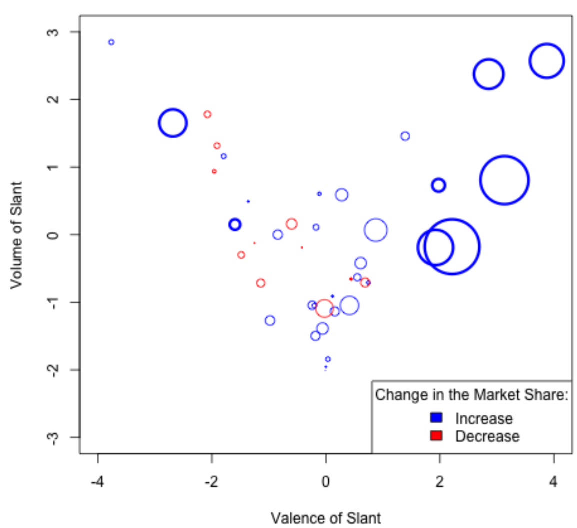
Figure 6: Predicted changes in the news outlets' market shares with the change in the amount of sensitive news by news outlet.



(a) Volume of internally sensitive news reporting



(b) Volume of Ukraine-crisis news reporting



(c) Slant in Ukraine-crisis news reporting

Each point represents a news outlet. The size of the points represents the degree of change of the market share of news outlets, measured as a percent of average market shares of this news outlet. The blue color corresponds to the increase in the market shares, and the red color corresponds to the decrease in the market share. The bold borders of the points correspond to significance of the change in the market share.

results of this regression based on b_j^{IS} with the different level of controls in regression (1), $\hat{b}_j^{IS} = d_0^{IS} + \bar{F}_j^{IS} d_1^{IS} + \zeta_j^{IS}$. In the specification with weekday and week fixed effect (column 4) that we've used above, the relationship between $\bar{F}_j^{IS}$ and $\hat{b}_j^{IS}$ is on the margin of being significant ($p < .05018$). In the three other specifications of regression (1) that are less restrictive (columns 1-3), the relationship between $\bar{F}_j^{IS}$ and $\hat{b}_j^{IS}$ is significant either on 5% or is on the margin of significance. We interpret this as evidence that news outlets with more reporting about the internally sensitive news are more likely to get an increase in their market shares on the day with more sensitive news, suggesting that an average consumer prefers internally sensitive news to non-sensitive news.

Table 6: Relationship between the estimates of the market share changes of news outlets, b_j^{IS}, and their ideological positions on internally sensitive news $\bar{F}_j^{IS}$.

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

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

Subfigures (b) and (c) of Figure 6 visualize the estimates of b_j^{Ukr}. Results suggest that news outlets with more reporting about the Ukraine-crisis news (subfigure b), lower pro-government valence of slant and higher volume of slant (subfigure c) are more likely to get an increase in the market shares on the days with more news about the Ukraine crisis. Similar to the case above, we test this relationship more formally by regressing the b_j^{Ukr} estimates on the average share of reporting about the Ukraine-crisis news, $\bar{F}_j^{Ukr}$, valence of slant in the reporting, val_j, and volume of slant, vol_j. Table 7 presents the the results of this regression based on b_j^{Ukr} with the different level of controls in regression (1), $\hat{b}_j^{Ukr} = d_0^{Ukr} + \bar{F}_j^{Ukr} d_1^{Ukr} + \text{val}_j d_2^{Ukr} + \text{vol}_j d_3^{Ukr} + \zeta_j^{Ukr}$. Based on the specification with weekday and week fixed effect (column 4) that we've used above, there is statistically significant positive relationship between $\hat{b}_j^{Ukr}$ and $\bar{F}_j^{Ukr}$, val_j and vol_j, supporting the claim that the news outlets that report more about Ukraine crisis and contain less pro-government propaganda and more slant overall are more likely to gain higher market shares during the days with a lot of news about the Ukraine crisis. However, the relationships between $\hat{b}_j^{Ukr}$ and volume and valence of slant is more noisy in other specifications (columns 1-3).

Table 7: Relationship between the estimates of the market share changes of news outlets, b_j^{Ukr}, and their ideological positions on Ukraine-crisis news, $\bar{F}_j^{Ukr}$, val_j and vol_j.

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

Controls are included in the regression (1) estimating b_j^{Ukr}. Standard errors are heteroskedasticity consistent.

We need to be careful with the interpretation of the results above. On the one hand, the relationship between the market shares and the relative importance of sensitive news is likely to be causal, as it only requires the conditional independence assumption (CIA) to hold: $\log(\text{share})_{jt} \perp log(F_t^l)|Z_{jt}\ \forall j, l = \{IS, Ukr\}$. CIA is a plausible assumption given that $log(F_t^l)$ is determined by the number of sensitive news events that happen on day t, a process that is not controlled by the market participants.[56]

However, causal shifts in the market shares due to changes in the amount of sensitive news events does not necessarily translate to the corresponding consumer preferences. While the results suggest that an average consumer prefers internally sensitive (Subfigure a of Figure 6) and Ukraine crisis (Subfigure b) news to non-sensitive news, and Ukraine-crisis news with less pro-government slant and higher volume of slant (Subfigure c), changes in the market shares can be driven by other factors. For example, if some consumers prefer pro-government slant and other consumers prefer anti-government slant in the Ukraine-crisis coverage, the market shares of the heavily slanted outlets will increase more on the days with

[56] This 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 eastern Ukraine, etc.

a lot of news about Ukraine, but not because of consumer preferences for the volume of slant. Alternatively, maybe the market is full of conscientious consumers, and the heavily slanted outlets benefit from their consumption. Another explanations is that consumers who prefer internally sensitive news also have high persistent preferences for the independent outlets, so the market share of these outlets increase on the days with a lot of sensitive news because of consumer sorting.

These alternative explanations for changes in the market share motivate us to estimate a structural demand model that captures consumer heterogeneity and potential conscientious consumption behavior.

5 Empirical Specification

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

5.1 Empirical model

There are three types of news that happen every day: non-sensitive news, internally sensitive news and Ukraine-crisis news. The news event realizations are driven by a stochastic process that is not controlled by the market participants. We do not observe the news event realizations but observe the relative importance of the sensitive news topic over time, F_t^{IS} and F_t^{Ukr}, defined in Section 4.1. For estimation purposes, we normalize F_t^{IS} and F_t^{Ukr} to have a unit mean.[57]

There are J news outlets in the market. Each news outlet reports non-sensitive news, internally sensitive news and Ukraine-crisis news. The news outlets report a share of sensitive news that happen each day, $\bar{F}_j^{IS}$ and $\bar{F}_j^{Ukr}$, and have stable ideological positions in their slant reporting, val_j and vol_j, defined in Section 4.1. We normalize the reporting and slant positions of the news outlets to have a zero mean and unit standard deviations. The outlets choose their reporting and slant positions, but the controlled outlets face additional costs of reporting internally sensitive news and having pro-Ukraine slant. The outlets also differ in their persistent characteristics, summarized by α_j.

There are I consumers in the market. We assume that consumers are in the market for online news on the days when they are browsing online. On each consumption occasion τ

[57]Notice that it only equates the average amount of sensitive news across days, but we can still interpret $F_t^{IS} = F_t^{Ukr} = 0$ as a day t with no sensitive news.

on day t, consumer i can choose one news outlet or an outside option of not consuming any news.[58] 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. Thus, consumer i can visit news outlet j on day t at most once.[59]

On each day t, consumer i can have multiple news consumptions $T_{it} = \{1, \ldots, J+1\}$. Consumers choose outlets sequentially on choice occasions $\tau = \{1, \ldots, T_{it}\}$, where in the last choice occasion T_{it}, she chooses the outside option. For simplicity, we treat the number of choice occasions T_{it} as exogenous.[60]

At each choice occasion τ on day t, a consumer chooses an outlet j such that $u_{ijt\tau} \geq u_{ij't\tau}\ \forall j' \in \{0, \ldots, J\} : j' \neq j$. We denote consumers' choices as y. Adapting consumer utility defined in Section 2.2 to our empirical context, we get

$$u_{ijt\tau} = \alpha_{ij} + F_t^{Ukr}\left(\omega_i^{Ukr} + \bar{F}_j^{Ukr}\beta_i^{Ukr} + \gamma_i^{\text{val}}\text{val}_j + \gamma_i^{\text{vol}}\text{vol}_j + |\text{val}_j - s_{i\tau}^{\text{val}}|(\tau > 1)\rho_i\right) + \\ + F_t^{IS}\left(\omega_i^{IS} + \bar{F}_j^{IS}\beta_i^{IS}\right) + |\text{val}_j - s_{i\tau}^{\text{val}}|(\tau > 1)\eta_i + \text{state}_{it\tau}\pi_i + \epsilon_{ijt\tau}. \tag{2}$$

Parameters $\{\alpha_{ij}, \beta_i^{Ukr}, \beta_i^{IS}, \gamma_i^{\text{val}}, \gamma_i^{\text{vol}}, \rho_i\}$ are the consumer preferences of interest defined in 2.2. There are two deviations in our empirical utility specification from the stylized model. First, given that the news outlets hold stable ideological positions, we separate out $\{F_t^{Ukr}, F_t^{IS}\}$, measures of relative importance of sensitive news, and $\{\bar{F}_j^{Ukr}, \bar{F}_j^{IS}\}$, measures of the ideological positions of the news outlets in consumer utility. The coefficients $\{\omega_i^{Ukr}, \omega_i^{IS}\}$ are the reduced-form coefficients that capture the relative preference of the consumer i for reading a news outlet with average sensitive news reporting, valence and volume of slant on the days with more sensitive news.[61] Second, $\text{state}_{it\tau}$ captures whether consumer i has already visited j on day t, so it ensures that consumers never visit the same news outlet twice on day t.

We note that there are multiple assumptions underlying this empirical model. First, we assume that consumers know the relative importance of news topics on day t, F_t^{IS} and F_t^{Ukr}. We believe that this is a reasonable assumption since we define consumption as visits of news articles, before which consumers usually have exposure to some proxy of the overall

[58]Following Gentzkow and Shapiro (2015), we restrict consumer choice to at most one news outlet per consumption occasion because it is impractical for people to read multiple news articles at the same time. Our set-up does not restrict consumers to navigate to multiple news outlets on the same day t.

[59]This discrete-choice specification ignores the intensity of news consumption within the outlet but significantly simplifies the computationally-intensive estimation process.

[60]This assumption limits our ability to simulate changes in the number of choice occasions in the counterfactual scenarios but still allows us to estimate consumer preferences for the variety in the news outlets' ideology.

[61]As we have normalized the reporting and slant to mean zero.

set of topics that have happened on day t, informing them about F_t^{IS} and F_t^{Ukr}. Second, we assume that consumers know the reporting and ideological positions of the news outlets, $\bar{F}_j^{IS}$, $\bar{F}_j^{Ukr}$, val_j and vol_j. In our estimation, we focus only on frequent news consumers, who are more likely to know the average reporting positions. Third, we assume that consumers do not incur switching costs when making day-to-day outlet consumption decisions. Any inertia of consumer choices or accumulated outlet brand capital (Bronnenberg and Dubé, 2016) is captured by the persistent preference coefficient, α_{ij}. Finally, following the stylized model in Section 2, we assume that consumer preferences are stable over time and that the news outlets differ in their reporting across all news topics and not within some particular news topics.

Identification of consumer preferences $\{\alpha_{ij}, \beta_i^{Ukr}, \beta_i^{IS}, \gamma_i^{\text{val}}, \gamma_i^{\text{vol}}, \rho_i\}$ relies on the exogenous shifts in F_t^{Ukr} and F_t^{IS}, the reporting and ideological positions of the news outlets $\bar{F}_j^{IS}$, $\bar{F}_j^{Ukr}$, val_j and vol_j, and across- and within-day news consumption choices, y.

5.2 Estimation Sample

In our model estimation, we focus on the news readers who consume news at least 10 days in our data sample period. These consumers are more likely to satisfy the assumptions underlying the model, such as the knowledge of the average news outlet's reporting and ideological positions, $\bar{F}_j^{IS}$, $\bar{F}_j^{Ukr}$, val_j and vol_j, and relative importance of sensitive news over time, F_t^{Ukr} and F_t^{IS}. There are 52,568 such news consumers in our sample.[62] We also focus on the top 36 online news outlets in the sample since the rest of the news outlets have a few number of consumer choices.

News readers in the selected sample have 4,456,161 consumption occasions, or outlet-day visits. On the majority (63.9%) 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 2.84 news outlets.

5.3 Estimation

We estimate the distribution of $\theta_i = \{\alpha_{ij}, \beta_i^{Ukr}, \beta_i^{IS}, \gamma_i^{\text{val}}, \gamma_i^{\text{vol}}, \rho_i, \omega_i^{IS}, \omega_i^{Ukr}, \eta_i, \pi_i\}$ using a Bayesian hierarchical model. We make standard logit assumption on $\epsilon_{ijt\tau}$, meaning that it follows a type-1 extreme value distribution, but allow for a flexible heterogeneity in consumer

[62]Out 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.

preferences. The probability that consumer i chooses news outlet j at on day t on the consumption occasion τ is

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

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

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

We use a normal distribution on the first-stage prior of θ_i, a normal prior over its mean and an inverse Wishart prior over the covariance matrix:

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

$$\mu \sim N(\bar{\mu}, \Sigma \otimes a_\mu^{-1}),$$
$$\Sigma \sim IW(\nu_\Sigma, \Psi_\Sigma).$$ [63]

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 4.2. However, this comes at a high computational costs, making the MCMC hybrid sampling procedure memory- and time-intensive. Online Appendix 9.8 provides more details about the sampling procedure.

6 Results and Counterfactuals

6.1 Consumer Preference Estimates

Table 8 presents the posterior point estimates of the persistent preferences of consumers, α_j, across the types of the 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, $\hat{\bar{\alpha}}$, to make the estimates more comparable across the news outlet types. The estimates reveal that an average consumer has the highest persistent preference for the GC news outlets ($\hat{\alpha}_{GC} - \hat{\bar{\alpha}} = 0.606$), followed by the potentially influenced ($\hat{\alpha}_{Inf} - \hat{\bar{\alpha}} = 0.312$) and independent ($\hat{\alpha}_{Inf} - \hat{\bar{\alpha}} = -0.072$) news outlets. While

[63] We set standard tuning parameters following Rossi et al. (2005) and Rossi (2014). Given the amount of data in our likelihood function, the results are almost unaffected by changing the tuning parameters.

there is substantial heterogeneity in consumer preferences, the vast majority of consumers have higher persistent preferences for the GC and potentially influenced news outlets than for an average news outlet (88.5% and 82.48%, respectively). These results imply that persistent preferences are at least one of the sources of demand for the GC news outlets.

Table 8: Posterior point estimates of persistent preferences for news outlets.

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

The posterior standard deviation estimates are in the brackets.

Table 9 reports the estimates of consumer preferences for sensitive news and ideological slant. An average news consumer has a relative preference for the internally sensitive ($E(\hat{\beta}^{IS}) = 0.021$) and Ukraine-crisis ($E(\hat{\beta}^{Ukr}) = 0.101$) news compared to non-sensitive news and prefers the Ukraine-crisis news with less pro-Russia valence ($E(\hat{\gamma}^{\text{val}}) = 0.068$) and less volume ($E(\hat{\gamma}^{\text{vol}}) = -0.01$) of slant. Thus, an average consumer has a distaste for the pro-government bias in the news, both in terms of censorship of internally sensitive news and pro-Russia slant in the Ukraine-crisis news.

However, the estimates in Table 9 also reveal that there is a substantial heterogeneity in consumer preferences. In particular, there are 42.11% of consumers who prefer non-sensitive news to internally sensitive news and 38.54% of consumers who prefer the Ukraine-crisis news with more pro-Russia slant, indicating their preferences for the pro-government bias. While an average consumer dislikes the ideological positions of the GC news outlets, the demand for the GC news outlets might be driven by the segment of consumers whose preferences are aligned with the government's ideological position.

To understand whether the demand for the GC news outlets primarily comes from the persistent preferences of consumers or from the ideological preferences of a segment of pro-government supporters, we compare the magnitudes of the consumer preferences. Recall

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

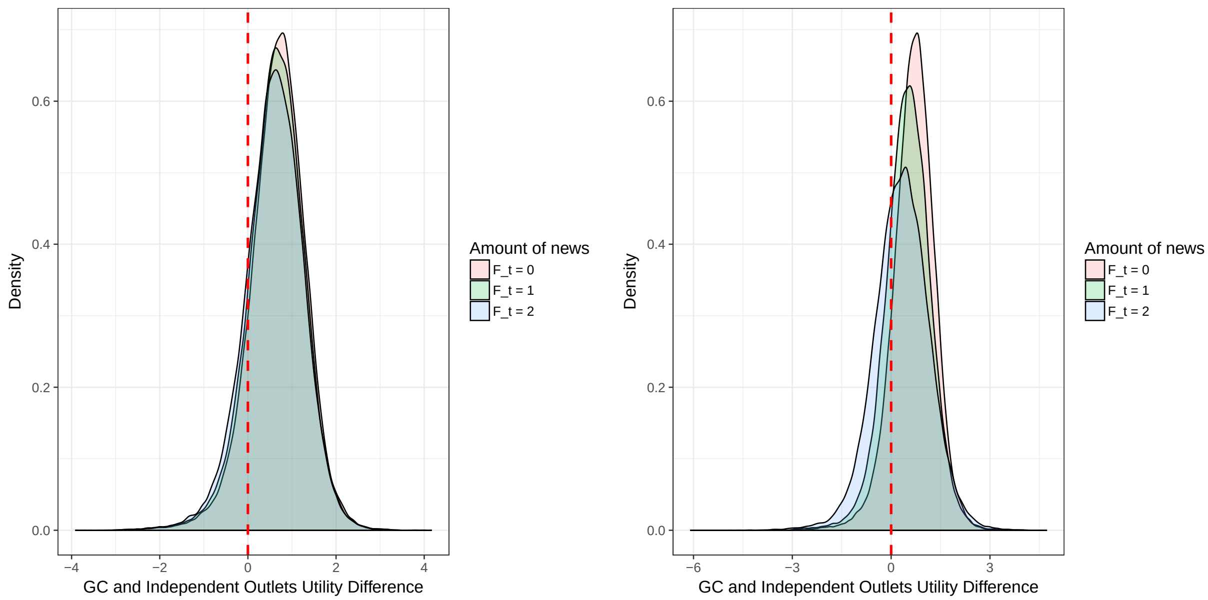
	Mean	S.D.	% of users > 0
$\hat{\omega}^{IS}$	0.001	0.135	50.41
	(0.001)	(0.001)	(0.36)
$\hat{\beta}^{IS}$	0.021	0.113	57.89
	(0.002)	(0.001)	(0.61)
$\hat{\omega}^{Ukr}$	0.262	0.568	68.56
	(0.005)	(0.003)	(0.31)
$\hat{\beta}^{Ukr}$	0.101	0.222	67.60
	(0.008)	(0.003)	(1.48)
$\hat{\gamma}^{\text{val}}$	0.068	0.235	61.46
	(0.003)	(0.002)	(0.53)
$\hat{\gamma}^{\text{vol}}$	-0.010	0.166	47.56
	(0.002)	(0.002)	(0.56)
$\hat{\eta}$	0.152	0.345	66.19
	(0.002)	(0.002)	(0.25)
$\hat{\rho}$	-0.109	0.164	25.50
	(0.002)	(0.001)	(0.36)

The posterior standard deviation estimates are in the brackets.

from Section 5.1 that we have normalized F_t^{IS} and F_t^{Ukr} to have a unit mean and $\bar{F}_j^{IS}$, $\bar{F}_j^{Ukr}$, val_j and vol_j to have a zero mean and unit standard deviation. This implies that $\hat{\omega}^{IS}$ and $\hat{\omega}^{Ukr}$ represent the extra utility of consuming a news outlet with average reporting and slant positions on a day with extra average amount of sensitive news, and $\hat{\beta}^{IS}$, $\hat{\beta}^{Ukr}$, $\hat{\gamma}^{\text{val}}$ and $\hat{\gamma}^{\text{vol}}$ represent the utility the consumer gets from one standard deviation more of reporting of internally sensitive news, Ukraine-crisis news, and valence and volume of slant in the Ukraine-crisis news.

Using these magnitudes, we can compute the difference in consumer utilities from the GC and independent news outlets coming from the controlled outlets' ideological positions. In the internally sensitive news, censorship is the method of government control, so the ideological difference is characterized by $F_t^{IS}\hat{\beta}_i^{IS}(\bar{F}_{GC}^{IS} - \bar{F}_{Ind}^{IS})$. The difference in coverage between the GC and independent news outlets is 2.17 standard deviations, so an average consumer gets $0.021 * 2.17 = 0.046$ more utility from an average independent news on the days with an average amount of internally sensitive news, normalized $F_t^{IS} = 1$. This utility difference pales compared to the 0.678 difference in persistent preferences for the GC and independent news outlets, suggesting that persistent preferences are more important in driving demand to the GC news outlets than the news coverage preferences. Subfigure (a) in Figure 7 confirms this by plotting changes in the utility difference consumers get from an average GC and independent outlets as there is more internally sensitive news, normalized $F_t^{IS} = \{0, 1, 2\}$. As F_t^{IS} increases, the fraction of consumers who prefer the GC news outlet stays almost the same, decreasing only from 78.4% to 73.8% as normalized F_t^{IS} changes from 0 to 2.

Figure 7: Distribution in the expected utility difference between an average GC and independent news outlet.



(a) Internally sensitive news

(b) Ukraine-crisis news

For the Ukraine-crisis news, ideological slant is the mechanism of government control, so the ideological difference between the GC and independent news outlets is characterized by $F_t^{Ukr}(\hat{\gamma}^{\text{val}}(\text{val}_{GC} - \text{val}_{Ind}) + \hat{\gamma}^{\text{vol}}(\text{vol}_{GC} - \text{vol}_{Ind}))$. There is a 2.45 standard deviation difference in the valence of slant ($\text{val}_{GC} - \text{val}_{Ind} = -2.45$) and 1.12 standard deviation difference in the volume of slant ($\text{vol}_{GC} - \text{vol}_{Ind} = 1.12$) between an average GC and independent news outlets, implying that an average consumer gets $0.068 * 2.45 + (-0.01) * -1.12 = 0.178$ less utility from an average GC on the days with an average amount of Ukraine-crisis news, normalized $F_t^{Ukr} = 1$. This utility difference is still smaller than the 0.678 difference in persistent preferences for the GC and independent news outlets. Subfigure (b) in Figure 7 shows that while an increase in the Ukraine-crisis news tilts extra consumers in favor of an average independent outlet, the majority (60.7%) of consumers still prefer an average GC outlet even on the days with a lot of Ukraine-crisis news, normalized $F_t^{Ukr} = 2$.

The results above show that persistent preferences are the main driver of the demand for the GC news outlets. In Section 6.2, we further examine the effect of the persistent preferences and the pro-government ideological positions on the market shares of the GC news outlets. However, before we move to the counterfactual simulations, we pause to discuss two features of the consumer preferences for news more deeply.

First, given the importance of the persistent preferences in consumers' demand for news outlets, it would be useful to understand the nature of these preferences. In particular, persistent preferences α_{ij} can represent some fixed characteristics of the news outlets, such as overall quality of the website or particular subset of the news topics that it covers, or some accumulated brand preferences of consumers that can change over time. These underlying mechanisms have different implications for our understanding of the level of government control over the news consumers. If α_{ij} represents only some fixed characteristics, such as quality of the website, government has full control over the news reporting of the GC outlets because this reporting does not affect persistent preferences. In contrast, if α_{ij} represents some accumulated brand capital of the website, the GC outlets cannot adopt viewpoints on the sensitive news events that are too extreme since consumers might stop reading the GC news outlets and change their persistent preferences in the long run. Unfortunately, we do not have a clean identification strategy to disentangle these potential mechanisms behind α_{ij}. However, we can use the estimated correlations of α_{ij}, $\hat{\Sigma}$, to check if the ideological positions of the news outlets explain some correlation in consumers' persistent preferences. In particular, if consumers have a higher persistent preference α_{ji} for the two websites that are ideologically more similar, it would suggest that the ideology of the websites affects persistent preferences formation. Indeed, we find that ideologically-similar news outlets tend to have higher correlation in α_{ij}, as described in Appendix 8.1. We conclude that the estimates of consumer preferences for ideology, β and γ, capture only the short-term effect of the ideological positions of the news outlets on the market shares.

Second, structural estimates allow us to examine the mechanism behind consumer preferences for the ideological slant. The results in Table 9 reveal that only a minority of consumers in the sample, 25.5%, behave like the conscientious news readers and consume a higher variety of slanted news on the days with a lot of sensitive news. Results also show an average preference of consumers for less volume of slant in the Ukraine-crisis news, although there is a substantial heterogeneity in consumer preferences. Both of these results support the theory that consumers prefer slanted sensitive news primarily because of the preference for like-minded news.

6.2 Counterfactuals

Consumer preference estimates reveal that persistent preferences drive consumer demand for the GC news outlets and that an average consumer prefers the ideological position of the independent news outlets to the GC outlets. These results suggest that the GC and potentially influenced news outlets get lower market shares because of government influence. What is the "cost" of the government control for these groups of news outlets? And, if we interpret the high persistent preferences of consumers for the GC news outlets as a result of government's investments, how much do the GC news outlets "benefit" in terms of their market share from these investments?

To answer these questions, we simulate the counterfactual scenarios in which we adjust the ideological positions and average persistent preferences for the GC news outlet. In these simulations, we treat the average ideological position of the independent news outlets as "unbiased," and interpret deviations from this ideological position as a result of government control. We interpret the results as the short-term reactions of the market to changes in the level of government control that do not account for the potential long-term reactions on the supply-side, such as product differentiation, and on the demand side, such as changes in persistent preferences.[64] In order the speed up the counterfactual simulation, we approximate the news realizations F_t^{IS} and F_t^{Ukr} by the centers of 20 clusters of these variables and simulate one choice occasion per consumer per day.[65]

Table 10 presents the simulated market shares under different levels of government control and persistent preferences of the GC news outlets. Columns (1) and (2) compare the predicted market shares under the current level of government control and a counterfactual scenario of no direct (ownership) government control (Gehlbach and Sonin, 2014), or when the GC news outlets have the same ideological positions as the independent news outlets, $\bar{F}_j^{IS,\text{new}} = \bar{F}_j^{IS} - \frac{\sum_{j' \in GC} \bar{F}_{j'}^{IS}}{\sum_{j' \in GC} 1} + \frac{\sum_{j' \in Ind} \bar{F}_{j'}^{IS}}{\sum_{j' \in Ind} 1}$, $\text{val}_j^{\text{new}} = \text{val}_j - \frac{\sum_{j' \in GC} \text{val}_{j'}}{\sum_{j' \in GC} 1} + \frac{\sum_{j' \in Ind} \text{val}_{j'}}{\sum_{j' \in Ind} 1}$ and $\text{vol}_j^{\text{new}} = \text{vol}_j - \frac{\sum_{j' \in GC} \text{vol}_{j'}}{\sum_{j' \in GC} 1} + \frac{\sum_{j' \in Ind} \text{vol}_{j'}}{\sum_{j' \in Ind} 1}$ $\forall\, j \in GC$. Without the direct control, market shares of the GC news outlets increase by 1.45 percentage points (20.2%), with most of the traffic coming from the extensive margin. Doing an approximate back-of the-envelope calculation, these 1.45 percentage points at most correspond to \$18.41 million of the advertising revenue.[66] For comparison, government subsidies to mass media in Russia in 2015 were \$1.21

[64] We leave a full equilibrium model of news supply under the government control constraints for future work.

[65] Standard k-means clustering algorithm is applied to cluster the observed F_t^{IS} and F_t^{Ukr}.

[66] While we do not have detailed information about the revenue sources of the news outlets, we can do a simple back-of-the-envelope calculation based on the advertising market size. For the online news outlets in

Table 10: Simulated market shares for different levels of government control and persistent preferences for the GC news outlets.

Outlet Type	(1)	(2)	(3)	(4)	(5)	(6)
Outlet Type	Actual α_{GC}		No control		More control	Low
(Market shares, %)		Direct	Indirect	Both		
GC ($\bar{\text{share}}_{Gov}$)	7.19	8.64	7.11	8.47	7.23	3.98
	(0.007)	(0.032)	(0.007)	(0.030)	(0.008)	(0.004)
Influenced ($\bar{\text{share}}_{Inf}$)	9.91	9.66	10.60	10.27	9.97	10.49
	(0.008)	(0.009)	(0.019)	(0.017)	(0.009)	(0.009)
Independent ($\bar{\text{share}}_{Ind}$)	6.35	6.19	6.25	6.12	5.92	6.60
	(0.009)	(0.009)	(0.009)	(0.010)	(0.017)	(0.010)
International ($\bar{\text{share}}_{Int}$)	0.63	0.61	0.62	0.60	0.63	0.66
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Ukrainian ($\bar{\text{share}}_{Ukr}$)	0.97	0.96	0.96	0.96	0.97	1.00
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
None above ($\bar{\text{share}}_{\text{Outside}}$)	74.95	73.94	74.46	73.59	75.28	77.27
	(0.014)	(0.033)	(0.024)	(0.045)	(0.016)	(0.014)

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

billion, which is around 65.7 times higher than the advertising loss.[67] Thus, it is relatively inexpensive for the government to "reimburse" the potential advertising losses the GC news outlets.

Similarly, we compute the expected advertising loss from the indirect control (Gehlbach and Sonin, 2014) or of the ideological positions of the potentially influenced news outlets. Column (3) presents the market shares under no indirect control, or when $\bar{F}_j^{IS,\text{new}} = \bar{F}_j^{IS} - \frac{\sum_{j'\in Inf} \bar{F}_{j'}^{IS}}{\sum_{j'\in Inf} 1} + \frac{\sum_{j'\in Ind} \bar{F}_{j'}^{IS}}{\sum_{j'\in Ind} 1}$, $\text{val}_j^{\text{new}} = \text{val}_j - \frac{\sum_{j'\in Inf} \text{val}_{j'}}{\sum_{j'\in Inf} 1} + \frac{\sum_{j'\in Ind} \text{val}_{j'}}{\sum_{j'\in Ind} 1}$ and $\text{vol}_j^{\text{new}} = \text{vol}_j - \frac{\sum_{j'\in Inf} \text{vol}_{j'}}{\sum_{j'\in Inf} 1} + \frac{\sum_{j'\in Ind} \text{vol}_{j'}}{\sum_{j'\in Ind} 1}$ $\forall\, j \in Inf$. Without the indirect control, the potentially influenced news outlets would have 0.94 percentage points (6.5%) higher market shares, corresponding to an upper bound of \$11.9 million.

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

Russia in 2013-2015, the main source of revenue is display advertising. In 2014, the total expenditure on display advertising on the Russian internet was 19.1 billion rubles (`http://www.akarussia.ru/knowledge/market_size`), which is around \$318 million using the exchange rate of the end of 2014 of 60 rubles for a dollar. Even if we assume that the online news market gets all the display advertising revenues, the 1.45 percentage points reduction in the market share of the GC news outlets due to direct control is small, corresponding to $\frac{1.45}{25.05} * \$318 = \$18.41$ million.

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

outlets increase slightly less as we remove control.[68]

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, $\bar{F}_j^{IS,\text{new}} = \bar{F}_j^{IS} - \frac{\sum_{j' \in Ind} \bar{F}_{j'}^{IS}}{\sum_{j' \in Ind} 1} + \frac{\sum_{j' \in Inf} \bar{F}_{j'}^{IS}}{\sum_{j' \in Inf} 1}$, $\text{val}_j^{\text{new}} = \text{val}_j - \frac{\sum_{j' \in Ind} \text{val}_{j'}}{\sum_{j' \in Ind} 1} + \frac{\sum_{j' \in Inf} \text{val}_{j'}}{\sum_{j' \in Inf} 1}$ and $\text{vol}_j^{\text{new}} = \text{vol}_j - \frac{\sum_{j' \in Ind} \text{vol}_{j'}}{\sum_{j' \in Ind} 1} + \frac{\sum_{j' \in Inf} \text{vol}_{j'}}{\sum_{j' \in Inf} 1}$ $\forall\ j \in Ind$, a feasible scenario given the events of 2016-2017.[69] In this case, independent news outlets lose 6.8% of their market share, corresponding to an upper bound of \$5.46 million. Once again, these results imply that it is not expensive for the government to convince the independent news outlets to become influenced, if the independent outlets care only about advertising revenues.

Finally, in column (6) we examine the scenario when the GC news outlets have lower average persistent preferences, such that they match the persistent preferences of consumers for the independent outlets, $\alpha_j^{\text{low}} = \alpha_j - \frac{\sum_{j' \in GC} \alpha_{j'}}{\sum_{j' \in GC} 1} + \frac{\sum_{j' \in Ind} \alpha_{j'}}{\sum_{j' \in Ind} 1}$ $\forall\ j \in GC$. Under the lower persistent preference regime, the market share of the GC news outlets decreases by 3.21 percentage points, or 44.6%.[70] If higher persistent preferences of the GC news outlets are driven by the government's investments in the outlets' quality (such as a better website, broader set of topics covered, etc.), these results imply that the government can almost double the market shares of the controlled outlets by the heavy investments in quality. In the Appendix 8.2, we show that this higher persistent preferences translate to 11.12 extra percentage points of the attention share of consumers, increasing the "media power" (Prat, 2017) of the GC news outlets from 0.28 to 0.5, or, in other words, it allows the GC news outlets to swing 25-75% elections instead of 36-64% elections in their favor.[71] These results hold with a similar magnitude for the days with a lot of sensitive news. We label the extra media power coming from higher persistent preferences as "brand media power," and conclude that it plays an important role in the Russian online news market.

[68]This is expected since we place the GC and potentially influenced outlets in the similar ideological positions.

[69]By the middle of 2016, several independent news outlets had to change their ownership due to a new law (`https://rg.ru/2016/01/01/smi-site-anons.html`), and rbc, one of the top online news outlets in Russia, had to change the editorial team due to the government pressure (`http://www.bbc.com/russian/news/2016/05/160513_rbc_badanin`) as well as its ownership later in 2017 (`http://www.forbes.ru/milliardery/346333-berezkin-kupil-u-prohorova-rbk`).

[70]This decrease correspond to an upper bound of \$40.7 million in advertising revenues.

[71]This holds under the worst-case scenario of naive news readers naive who do not understand that the GC news outlets are trying to persuade them. For more details, see Prat (2017) and Kennedy and Prat (2017).

7 Conclusion

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 persistent preferences of consumers for the outlets. This implies that the government can impose its ideological position on the news readers, at least in the short run, by heavily investing in factors that increase the persistent preferences of consumers for the news outlets, such as outlets' websites or news product quality. In the Russian online news market, we find that the high persistent preferences for the GC news outlets are responsible for 44.6% of their market share, while the pro-government ideological bias in the news reduce the GC outlets' market share by 20.2%, suggesting that the Russian government controls the ideological diet of a large group of the online news consumers.

We note that our results should be taken with two caveats in mind. First, the ideological preferences of consumers in our sample might not extrapolate to the entire population in Russia. Given the nature of our empirical exercise, we focus only on the online news consumers, who might be different from an average news consumer in Russia who cites TV as the main source of news.[72] Indeed, most political surveys have indicated the overwhelming support of the government (e.g. in the Ukraine-crisis handling) during the period of our study,[73] and it is unclear whether our estimates differ because of a bias in the stated preferences in the surveys or because of selection on the ideological preferences to consuming news online. However, we consider the ideological preferences of the online news consumers important on their own, since these preferences define the ability of the governments to control the news online, which is a growing segment of news consumption.[74]

Another feature of consumers in our sample is that they use the Internet Explorer Toolbar. IE users tend to be older, more work-oriented, and perhaps less technologically-savvy than an average news consumer in Russia. While in Section 3.3.1 we show that the aggregate patterns of news consumption of IE Toolbar users match the overall news consumption in Russia, such aggregate similarity can hide differences in the distributions of ideological preferences. However, even if there are ideological differences between the IE Toolbar users

[72] `https://wciom.ru/index.php?id=236&uid=116190` (Russian).

[73] `https://www.economist.com/blogs/graphicdetail/2016/02/daily-chart-4`.

[74] `https://wciom.ru/index.php?id=236&uid=116190` (Russian); `http://www.pewresearch.org/fact-tank/2017/10/04/key-trends-in-social-and-digital-news-media/`.

and average news consumers in Russia, we would expect that the IE Toolbar users should be more pro-government than an average news reader, reinforcing the idea that consumers' ideological preferences are not the main driver behind the demand for the GC news outlets.[75]

Second, while persistent preferences of consumers play a large role in their news consumption, we need to be careful with interpretations. On the one hand, persistent preferences of consumers can represent preferences for some characteristics of the news outlets that are unrelated to the outlets' ideological positions, such as quality of the outlets' websites, news coverage, or other things. In this case, governments have a powerful tool of "promoting" even extreme ideological positions in the controlled news outlets, and all they need to do is invest in the important outlet characteristics. On the other hand, persistent preferences of consumers can also consist of the accumulated brand capital of the news outlets that can change over time. If this is the case, the government can "promote" their ideological position only in the short run, since in the long run consumers might change their perception of the controlled news outlets. In Section 6.1, we show that there is a relationship between in the ideological positions of the news outlets and the structure of persistent preferences. We leave the questions of the degree and speed of persistent preference adjustments for future research.

References

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*.

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.

[75]Intuitively, if there is selection in our estimates and given that the IE Toolbar users are older and less technologically-savvy, we would expect that they are more pro-government than an average online news reader in Russia, since older people tend to show higher support for the Russian government (`https://republic.ru/russia/kto_eti_lyudi_podderzhivayushchie_putina-745894.xhtml`, Russian).

Bronnenberg, B. J. and J.-P. Dubé (2016). The formation of consumer brand preferences. *Annual Review of Economics* (0).

Cagé, J., N. Hervé, and M.-L. Viaud (2017). The production of information in an online world: Is copy right? *Working Paper*.

Chevalier, J. A. and D. Mayzlin (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research 43*(3), 345–354.

Cohen, B. (1963). The press and foreign policy. *Institute of Governmental Studies, University of California*.

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.

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.

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.

Franceschelli, I. (2011). When the ink is gone: The transition from print to online editions. *Working Paper*.

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. 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 (2016). Measuring polarization in high-dimensional data: Method and application to congressional speech. *Working Paper, National Bureau of Economic Research*.

Groseclose, T. and J. Milyo (2005). A measure of media bias. *The Quarterly Journal of Economics 120*(4), 1191–1237.

Hotelling, H. (1929). Stability in competition. *The economic journal 39*(153), 41–57.

Iyengar, S. and D. R. Kinder (1987). News that matters: Agenda-setting and priming in a television age. *University of Chicago Press*.

Kennedy, P. and A. Prat (2017). Where do people get their news? *Working Paper*.

Kim, J., G. M. Allenby, and P. E. Rossi (2002). Modeling consumer demand for variety. *Marketing Science 21*(3), 229–250.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of learning and motivation 32*, 385–418.

Knight, B. and A. Tribin (2016). The limits of propaganda: Evidence from chavez's venezuela. *Working Paper, National Bureau of Economic Research*.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing 70*(3), 74–89.

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.

Mullainathan, S. and A. Shleifer (2005). The market for news. *The American Economic Review 95*(4), 1031–1053.

Pan, J. and Y. Xu (2017). China's ideological spectrum. *The Journal of Politics Forthcoming*.

Perego, J. and S. Yuksel (2016). Media competition and the source of disagreement. *Working Paper*.

Petrova, M. (2008). Inequality and media capture. *Journal of public Economics 92*(1), 183–212.

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*.

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.

Sen, A. and P. Yildirim (2016). Clicks bias in editorial decisions: How does popularity shape online news coverage? *Working paper*.

Thussu, D. K. (2008). *News as entertainment: The rise of global infotainment*. Sage.

Tirole, J. (1988). *The theory of industrial organization*. MIT press.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge university press.

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.

8 Appendices

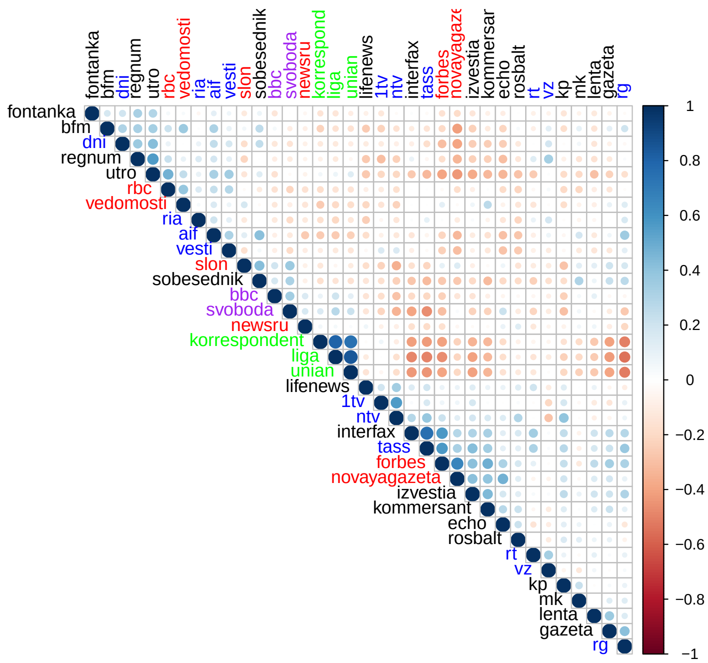
8.1 Appendix A: Correlation in Persistent Preferences

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

Figure 8 summarizes the estimates of correlation in persistent outlet preferences, α_{ij}, across the news outlets. Similar to Table 8, we subtract the average preference for news outlets, $\bar{\alpha}_i$, from the α_{ij} to exclude the influence of consumer i's preference for news in general. News outlets are colored by their types, corresponding to the legend in Figure 5, and are sorted by the degree of correlation between each other. The results suggest that there is at least some correlation in consumer persistent preferences driven by the news outlets' ideology. For example, consumer preferences for all Ukrainian and international news outlets are highly positively correlated among each other and are negatively correlated with the GC news outlets. At the same time, news outlets are not perfectly grouped by types, suggesting that other website characteristics might also play a role in the persistent preferences.

To test the alternative explanations for persistent preferences of consumers more formally, we regress the estimated correlations on the ideological and quality distance between the news outlets. We measure the distance as the absolute value in the news outlets characteristics,

Figure 8: Posterior estimates of the correlation matrix of persistent consumer preferences for news websites, $\alpha_{ij} - \bar{\alpha}_i$.



Each dot represents the correlation of $\alpha_j - \bar{\alpha}_j$ for two news outlets. The scale on the right explains the color code of the correlations. The colors of the text labels correspond to types of the news outlets used throughout the draft (first explained in Figure 5).

such as the amount of reporting about sensitive news, $\bar{F}_j^{IS}$ and $\bar{F}_j^{Ukr}$, valence and volume of slant about the Ukraine-crisis news, val_j and vol_j, and quality measured as $\bar{\alpha}_j$. To make the regression coefficients comparable, we normalize the standard deviation of the absolute value differences to 1. Table 11 presents the regression results. We can confirm that the ideological distance between the news outlets indeed has an effect on the correlations in the persistent preferences of the news outlets. For example, the news outlets that are 1 standard deviation more similar in the valence of slant in the Ukraine-crisis reporting tend to have 5.56% more correlated persistent preferences among the news consumers.

Table 11: Relationship between the correlations in persistent preferences of consumers, $\alpha_{ij} - \bar{\alpha}_i$, and distance between the outlets' characteristics.

	Dependent variable:
	$\text{cor}(\alpha_{ij} - \bar{\alpha}_i, \alpha_{ij'} - \bar{\alpha}_i)\ \forall\ j \neq j'$
Intercept	0.1155
	(0.0191)
$\lvert\bar{F}_j^{IS} - \bar{F}_{j'}^{IS}\rvert$	-0.0145
	(0.0076)
$\lvert\bar{F}_j^{Ukr} - \bar{F}_{j'}^{Ukr}\rvert$	0.0517
	(0.0076)
$\lvert\text{val}_j - \text{val}_j'\rvert$	-0.0556
	(0.0085)
$\lvert\text{vol}_j - \text{vol}_j'\rvert$	0.0037
	(0.0081)
$\lvert\bar{\alpha}_j - \bar{\alpha}_{j'}\rvert$	-0.0421
	(0.008)
Observations	630
R^2	0.2133
Adjusted R^2	0.207

The results above suggest that persistent preferences of consumers, α_{ij}, are correlated with the ideological positions of the news outlets, implying that in the long run persistent preferences of consumers might change as we change the ideological positions of the news outlets.

8.2 Appendix B: Online Media Power of the Government

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

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

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

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

Using this definition, the attention share of the GC news outlets is 33.4% (0.1%), corresponding to the media power of 0.5 under the worst-case scenario assumptions, meaning that the government is able to swing 25-75% elections into a draw.[76]

To understand the role of consumers' persistent preferences for the GC news outlets in the GC outlets' media power, we compute the attention shares of consumers under the lower persistent preferences for the GC news outlets, as in the case of column (6) of Table 10. Under these persistent preferences, the online attention share of the GC news outlets reduce by 11.12 percentage points to 21.78% (0.06%), corresponding to 0.28 media power. Such media power allows the government to swing 36-64% elections in to a draw. Thus, around 1/3 of the attention share of the GC news outlets and almost half of their media power is driven by high persistent preferences for the GC news outlets, which we refer to as "brand media power."

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

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

[76] The worst-case scenario includes the assumption that the readers are naive and do not understand that the GC news outlets are trying to persuade them. For more details, see Prat (2017) and Kennedy and Prat (2017).

where ΔU_i^x is the utility consumer i gets from the pro-government bias in sensitive news x topic, $\Delta U_i^{IS} = \hat{\beta}_i^{IS}(\bar{F}_{GC}^{IS} - \bar{F}_{Ind}^{IS})$ and $\Delta U_i^{Ukr} = \hat{\gamma}^{\text{val}}(\text{val}_{GC} - \text{val}_{Ind}) + \hat{\gamma}^{\text{vol}}(\text{vol}_{GC} - \text{vol}_{Ind})$. We use this measure to compute the attention share of the GC news outlets over consumers with $\Delta U_i^x < 0$ on the days with a lot of sensitive news, a case where the GC news outlets can successfully prevent a motivated consumer from learning the information. The attention shares are 31.1% for a big internally sensitive news day and 21.5% for a big Ukraine-crisis news day.[77] Under the lower persistent preferences for the GC news outlets, the attention shares on such days change to 19.9% and 13.1%, respectively. Thus, the high persistent preferences for the GC news outlets allows them to capture an additional 8.4-11.2 percentage points of consumers who prefer the ideological coverage of the independent news outlets.

[77] A big sensitive news day is a day with three times the average amount of sensitive news.

FOR ONLINE PUBLICATION

9 Online Appendices

9.1 News Outlets Classification

News outlets' classification is done based on the media ownership information, evidence of the indirect influence such as removing news articles because of the political pressure, and interviews with media professionals. Below we present more details about each group and news outlet.

Government-controlled news outlets:

- *vesti, 1tv, tass, rg, rt* and *ria* are directly owned by the government. Included as GC news outlets because of the ownership structure.

- *aif* is owned by Moscow City Hall. Included as a GC news outlet because of the ownership structure.

- *ntv* is owned by Gazprom, a state-owned gas monopolist. Included as a GC news outlet because of the ownership structure.

- *vz* and *dni* were founded by Konstantin Rykov, a member of an incumbent political party United Russia who led the political campaigns in support of Vladimir Putin in 2007. *vz* is owned by the Institute of Socio-Economics and Political Research, which is managed by Dmitry Badovsky, a former deputy chief of the Presidential Administration of Russia. Included as GC news outlets because of being founded by Rykov and managed by Badovsky.

Independent news outlets:

- *newsru* is owned by Vladimir Gusinsky, a tycoon who has opposed the incumbent Russian government since 2001. Included as an independent news outlet because of the ownership by Gusinsky and no reports of being influenced by the government.

- *newtimes* is owned by an investigative journalist Yevgenia Albats and a non-profit fund The New Times Foundation. Included as an independent news outlet because of the ownership by Albats and no reports of being influenced by the government.

- *novayagazeta* is owned by journalists (76%), Alexander Lebedev (14%) and Mikhail Gorbachev (10%). Included as an independent news outlet because of the ownership by journalists and no reports of being influenced by the government.

- *rbc* and *snob* are owned by Mikhail Prokhorov, a Russian billionaire and politician. He run for president in the 2012 elections. RBC.ru stayed independent until May 2016, when the top managers were fired due to political pressure.[78] Included as an independent news outlet because of the ownership by Prokhorov and no reports of being influenced by the government. It was later acquired by Grigory Berezkin, the owner of *kp.ru*, in June 2017.

- *slon* and *tvrain* are owned by Alexander Vinokurov and Natalia Sidneeva. *tvrain*'s TV channel was taken off the air by the major cable systems after covering 2011 street protests. Included as independent news outlets because of the ownership and no reports of being influenced by the government.

- *vedomosti* was jointly owned by Sanoma Independent Media (33%), Financial Times (33%) and The Wall Street Journal (33%) until the end of 2015. It was sold to Demyan Kudryavsev in November 2015 due to the a new law limiting foreign ownership of media to 20% starting in 2016. Included as an independent news outlet because of the international ownership structure and no reports of being influenced by the government.

- *forbes* was owned by Axel Springer before the end of 2015. It was sold to Alexander Fedotov in October 2015 due to a new law limiting foreign ownership of media to 20% starting in 2016. Included as an independent news outlet because of the international ownership structure and no reports of being influenced by the government.

- *the-village* is owned by Look at Media publishing, founded an managed by journalists Vasily Esmanov and Alexey Ametov. Included as an independent news outlet because of the being managed by journalists and no reports of being influenced by the government.

Potentially influenced news outlets:

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

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

in March of 2014.[79]. Included as potentially influenced news outlets because of the editorial changes.

- *izvestia* is owned by Yuri Kovalchuk through the National Media Group (NMG). Yuri Kovalchuk is a close friend of Vladimir Putin. Included as a potentially influenced news outlet because of the ownership by Kovalchuk.

- *lifenews* is owned by Aram Gabrelyanov, a manager of the National Media Group (NMG).[80] Included as a potentially influenced news outlet because of the ownership by Gabrelyanov linked to Kovalchuk.

- *kommersant* is owned by Alisher Usmanov, one of the richest Russian oligarchs.[81] Included as a potentially influenced news outlet because of the ownership by Usmanov.

- *kp* is owned by Grigory Berezkin, who is on the board of directors of state-owned RZD.[82] Included as a potentially influenced news outlet because of the ownership by Berezkin.

- *bfm* is owned by Rumedia, a company of Russian steel tycoon Vladimir Lisin.[83] Included as a potentially influenced news outlet because of the ownership by Lisin.

- *echo* is jointly owned by journalists of echo (34%) and a state-owned gas monopolist Gazprom (66%). One of the most famous Russian independent media companies, it is reported to be influenced by the government and publish paid articles.[84] Included as a potentially influenced news outlet because of the ownership by Gazprom and paid articles reports.

- *mk* is owned by Pavel Gusev, a confidant of Vladimir Putin. There are examples of mk removing published articles about government-sensitive topics.[85] Included as a potentially influenced news outlet because of the ownership by Gusev and removal of the news articles.

[79] https://meduza.io/feature/2016/05/17/12-redaktsiy-za-pyat-let
[80] http://www.kommersant.ru/doc/2311510
[81] https://lenta.ru/lib/14164974/
[82] http://www.forbes.ru/profile/grigorii-berezkin
[83] https://en.wikipedia.org/wiki/Vladimir_Lisin
[84] https://tjournal.ru/p/media-denim
[85] http://www.rbc.ru/politics/27/12/2013/897386.shtml

- *ng* is owned by Konstantin Remchukov. It is reported to publish articles which are paid for by the government.[86] Included as a potentially influenced news outlet because of the paid articles reports.

- *regnum* is reported to have been purchased by Gazprom media.[87] It is reported to publish paid articles.[88] Included as a potentially influenced news outlet because of the ownership by Gazprom media and paid articles reports.

- *rosbalt, sobesednik* and *trud* are reported to publish paid articles.[89] Included as potentially influenced news outlets because of the paid articles reports.

- For six news outlets, *polit, utro, ridus, fontanka, interfax and znak*, we could not find strong evidence of either being influenced or independent. The media professionals that we have interviewed indicated that these news outlet could be influenced by the government. Because of this, and in order to be conservative in our independent news outlets classification, we include them as potentially influenced.

International News Outlets:

- *bbc* is the Russian version of BBC.

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

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

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

- *reuters* is the Russian version of Reuters.

Ukrainian news outlets:

- *korrespondent, liga* and *unian* are all based in Ukraine.

[86] http://theins.ru/politika/6015
[87] https://lenta.ru/news/2014/06/20/media/
[88] https://tjournal.ru/p/media-denim
[89] https://tjournal.ru/p/media-denim

9.2 Publication Records Collection and Processing

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

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

To find sensitive news and the corresponding media slant, we process the texts of the news articles by stemming all the words and removing punctuation and stop words. We define proper nouns in the text corpus as any word that frequently (more than 50% of times used in the corpus) starts with a capital letter in the text when it is not at the beginning of the sentence.[90]

[90] In doing this, we include the typical proper nouns but exclude words that are used as proper nouns rarely and only in a certain context.

9.3 Sensitive News: Censorship and Slant

9.3.1 Censored unigrams and bigrams

Tables 12 and 13 present 54 bigrams of the proper nouns that are underused by the GC news outlets. To define a set of censored bigrams, we exclude the bigrams related to the profession of journalism, such as names of journalists, media owners, news outlets, etc. We also exclude three common actors, Dmitry Medvedev, Ramzan Kadyrov and Alisher Usmanov, given that there is a lot of regular news about these actors. The resulting set of censored bigrams of the proper nouns contain 34 bigrams (marked bold in the tables 12 and 13).

Table 12: List of the top 54 bigrams of the proper nouns underused by the GC news outlets. Part 1.

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

Bigrams marked as bold are selected to define sensitive news.

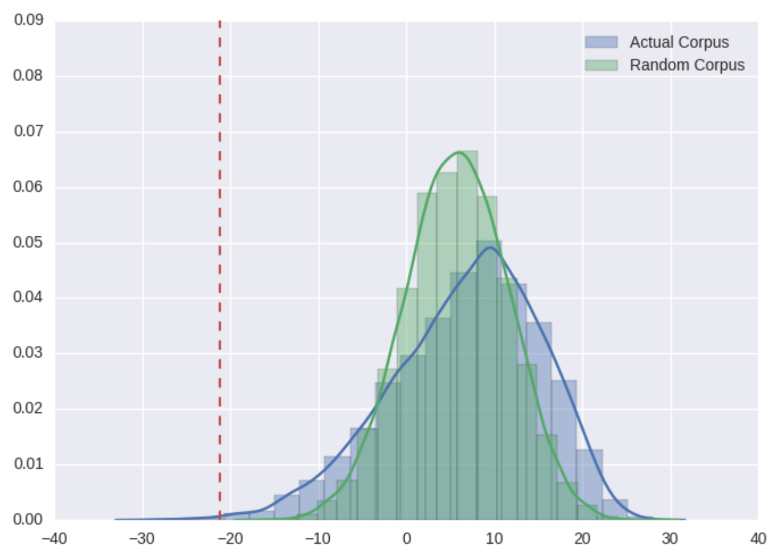
Table 13: List of the top 54 bigrams of the proper nouns underused by the GC news outlets. Part 2.

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

Bigrams marked as bold are selected to define sensitive news.

In addition to the bigrams of the proper nouns, we re-do the classification using the unigrams of the proper nouns. We do this to make sure that we do not exclude facts described with a single proper noun. Figure 9 presents the histograms of the rank difference distributions, $\Delta\text{Rank}_v^{\text{GC-Ind}}$ and $\Delta_k^{\text{Random}}\text{Rank}_v^{\text{GC-Ind}}$. To define censored proper nouns we compare the lowest rank difference in $\Delta\text{Rank}_v^{\text{GC-Ind}}$ (-29.3) and in $\Delta_k^{\text{Random}}\text{Rank}_v^{\text{GC-Ind}}$ (-21.1). There are 47 unigrams of the proper nouns in the actual sample with a rank difference below the threshold of -21.1. A lot of these unigrams correspond to the last names of the sensitive actors which are classified based on bigrams, and some others refer to the ambiguous actors.

Figure 9: Histograms of $\Delta\text{Rank}_v^{Ind-Gov}$ across the proper nouns: actual and random corpus.



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

Table 14 provides an example of the top 20 underused unigrams. To define a set of censored unigrams, we exclude the unigrams related to the profession of journalism, and unigrams that refer to ambiguous actors. The resulting set of censored unigrams contains 10 proper nouns (marked bold in the table 14).

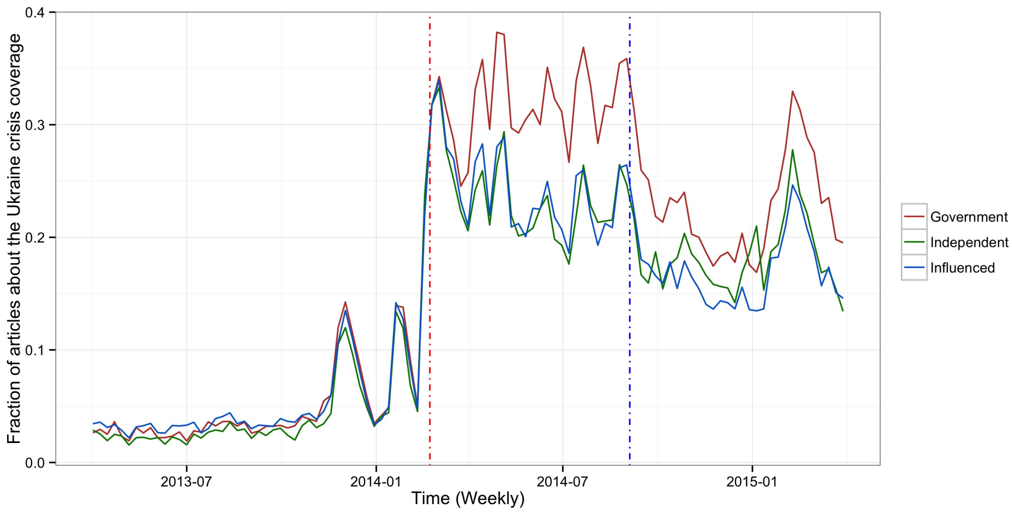
Table 14: List of the top 20 unigrams of the proper nouns underused by the GC news outlets.

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

Unigrams marked as bold are selected to define sensitive news.

9.4 Volume of News and Slant in the Ukraine-Crisis Reporting

Figure 10: Share of articles containing the word "Ukraine" in the weekly coverage of news outlets by types.



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

9.5 Summary of Browsing Behavior

Each news website consists of 4 different types of pages: the main page, news articles pages, news subdirectories, and other pages. We classify the visit as the main page visit if the visited URL matches the main page url. We classify the visit as the news article visit if the visited URL matches one of the URLs of the publication records data or has a structure similar to it.[92] We classify the URL as a subdirectory if the visited URL matches the subdirectory URL.[93] We classify the rest of the URL visits as other page visits. The majority of the URL visits classified as other pages correspond to the photos, videos and other special content on news websites.

[91] This is a Ukrainian word to describe protesters supporting the former Ukraine government.

[92] For example, if the article URL has the structure `http://www.x1.ru/news/topic/year/month/date/name-of-the-article.html`, we classify any ULR with the structure `http://www.x1.ru/news/topic/year/month/date/some-other-name-of-the-article.html` as news articles.

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

Table 15: List of the top 10 overused words by the GC and Ukrainian news outlets in the Ukraine-crisis news coverage.

Overused words by the:			
GC news outlets		Ukrainian news outlets	
Word	$\Delta\text{Rank}_v^{Ukr-Gov}$	Word	$\Delta\text{Rank}_v^{Ukr-Gov}$
reunion	34.7	continental	-31.7
radical	34.1	annexation	-30.8
punitive	33.5	monopolistic	-30
overturn	33.1	anti-terrorist	-29.9
blockade	32.6	devoid	-29.8
bombing	32.2	titushky[91]	-29.4
coup	31.7	content	-29.3
anti-Russian	31.1	residue	-29.3
colored	31	occupied	-29.3
deepest	31	deduced	-29.2

Table 16: List of the words corresponding to the pro-Russia and pro-Ukraine slant.

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

Classification is done using the top 50 words with the lowest and highest rank score differences $\Delta\text{Rank}_v^{\text{Ukr-GC}}$. We add three extra terms ranked 53, 60 and 64 to the pro-Russia slant since they represent another way of spelling of the pro-Russia slanted terms.

News articles account for most page views on news websites. Other webpages are visited half as often as news articles. The main directory and news subdirectories are also each visited only half as often as news articles. Table 4 shows statistics of browsing of the different webpage types. While some consumers read news from the headlines, most of the time the main pages and news subdirectories help readers to navigate to the news articles. This also includes navigation to the non-news content in the "other" sections. Thus, we only use navigation to news articles as records of news consumption.

9.6 Comparing Weekly Visitors of IE Toolbar and LI

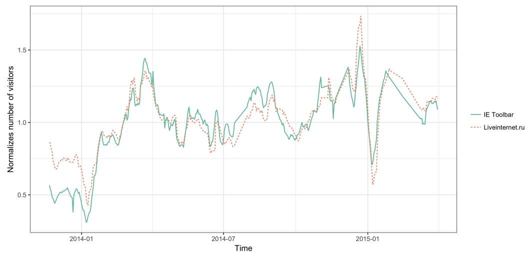
Table 17 presents the visit shares of the 23 out of the top 30 websites in the scraped LI data.[94] For the resulting set of websites, we collect usage information for the news readers in the IE Toolbar data.[95] IE Toolbar users are more likely to be older (visit odnoklassniki.ru, a social network with older demographics, more than vk.com, a social network with younger demographics), more interested in weather and less interested in streaming and entertainment websites. This is consistent with the anecdotes that IE Toolbar users are more likely to be accessing the internet from the office space. However, the shares and rankings of the website are relatively similar, suggesting that the IE Toolbar consumers are not too different from the general population in Russia.

Figure 11 presents the traffic to the top seven LI news outlets based on the LI and IE Toolbar data. For each website and news source, the average traffic level is normalized to one, and the IE Toolbar data are corrected for the attrition rate. Changes in the news consumption in the IE Toolbar data closely track the population-level consumption in the LI data, with the correlation between 0.52 and 0.914 for the news outlets. This high correlation between the IE Toolbar and LI news outlets' readership highlights that consumers who leave the IE Toolbar do not differ in their news consumption habits from those who stay in the sample. This is important given high attrition rates from the IE Toolbar data over the sample period, which is explained by the roll-out of new versions of the IE, Windows Edge browser, as well as consumers switching to other browsers.

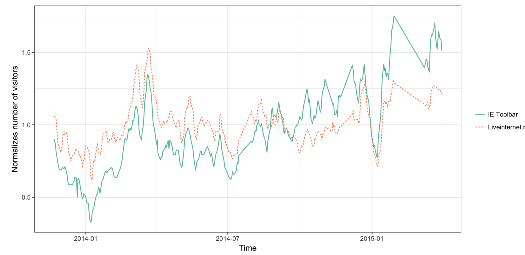
[94] We exclude subsections of mail.ru (e.g. auto.mail.ru) from the comparison since subsections of non-news websites were not extracted from the IE Toolbar data, as well as pulso.ru, a website that tracks clicks on the social media links on various websites, since it is not recorded in the IE Toolbar data.

[95] Unfortunately, we do not have information on the IE Toolbar users who are not the news readers. While our definition of the news readers is broad (visit a URL of the top 48 Russian online news outlets at least once over one and a half years), focusing only on browsing behavior of the news readers might lead to selection driving the differences between columns 3 and 4 of Table 17.

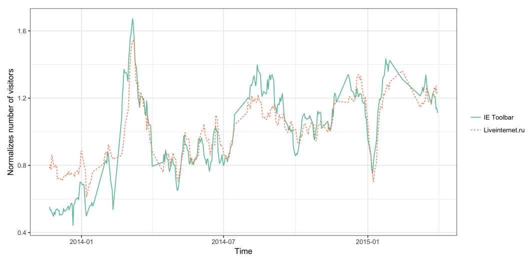
Figure 11: Normalized traffic of the top seven news websites, IE Toolbar and Liveinternet.ru



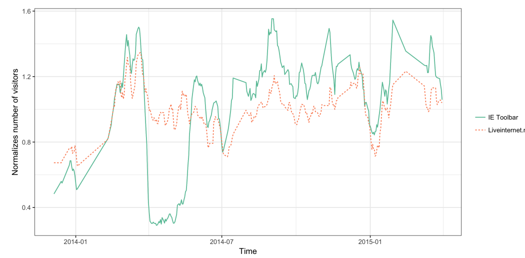
(a) ria.ru (cor = 0.914)



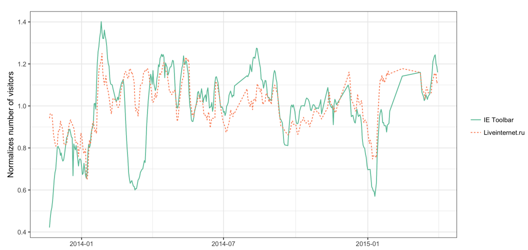
(b) ria.ru (cor = 0.702)



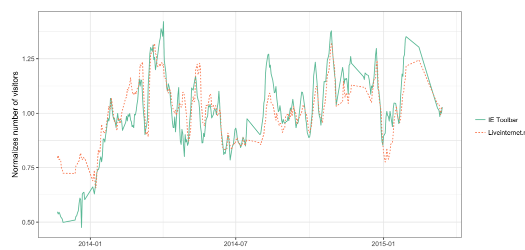
(c) lenta.ru (cor = 0.913)



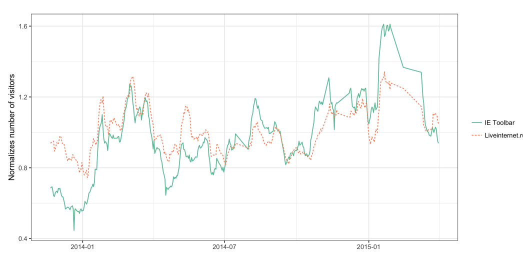
(d) gazeta.ru (cor = 0.520)



(e) vesti.ru (cor = 0.549)



(f) rg.ru (cor = 0.830)



(g) kp.ru (cor = 0.807)

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

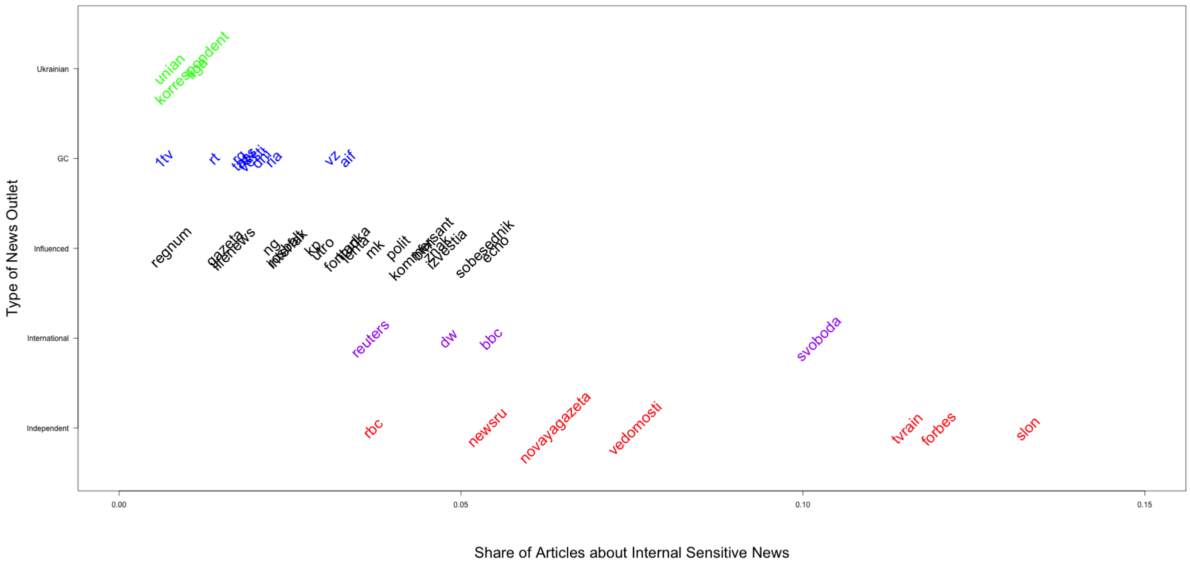
Table 17: Comparison of website rankings in IE Toolbar and LI.ru

Website	Description	Visit Share	
		IE Toolbar	liveinternet.ru
vk.com	Social Network (younger audience)	0.2904	0.3653
odnoklassniki.ru	Social Network (older audience)	0.3131	0.2522
avito.ru	Classified posts	0.0639	0.0682
gismeteo.ru	Weather	0.0489	0.0339
rbc.ru	News outlet	0.0458	0.0233
kinopoisk.ru	Movie descriptions	0.0096	0.0224
ria.ru	News outlet	0.0084	0.0214
vesti.ru	News outlet	0.0264	0.0183
rutracker.org	Torrent website	0.0051	0.0177
drom.ru	Website about cars	0.0193	0.0165
kp.ru	News outlet	0.0166	0.0159
lenta.ru	News outlet	0.0164	0.0159
gazeta.ru	News outlet	0.0156	0.0156
liveinternet.ru	Statistics tracking and blogging platform	0.0147	0.0151
ngs.ru	Novosibirsk city website	0.0060	0.0151
smi2.ru	News aggregation website	0.0360	0.0151
rg.ru	News outlet	0.0164	0.0149
zoomby.ru	Streaming website	0.0104	0.0143
auto.ru	Buy/Sell used cars	0.0115	0.0137
tiu.ru	Online retailer	0.0080	0.0136
hh.ru	Job postings	0.0146	0.0134
wildberries.ru	Online retailer	0.0130	0.0133
woman.ru	Online magazine	0.0074	0.0126

We remove pulso.ru, a tracker website that is not recorded by IE Toolbar data, and various version of mail.ru (e.g. auto.mail.ru) as the subsections of non-news websites were not extracted from the IE Toolbar data.

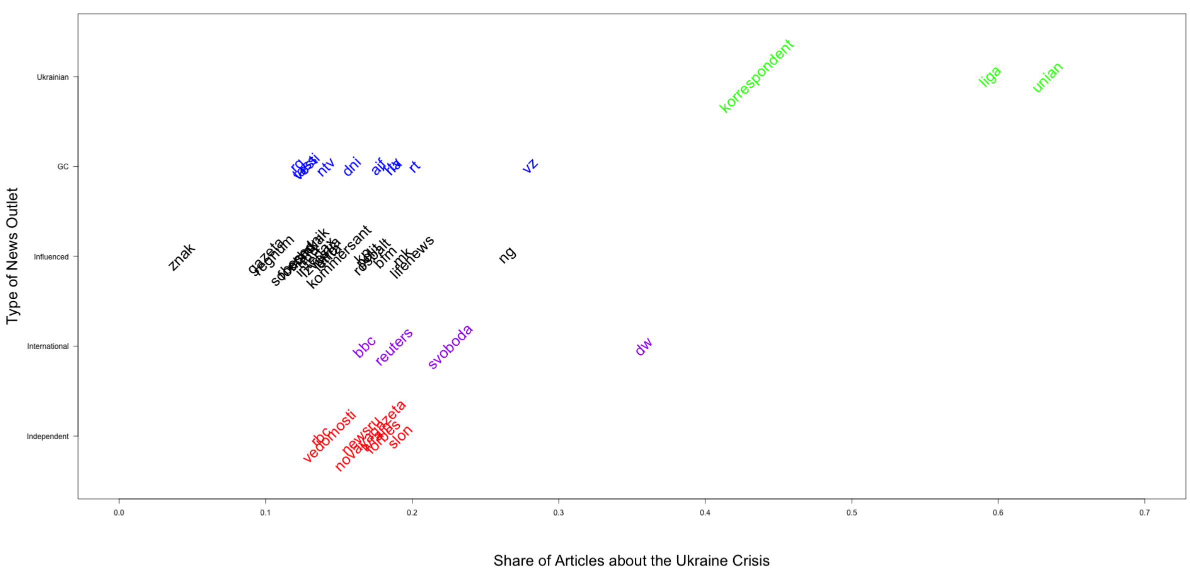
9.7 Ideological Positions of the News Outlets

Figure 12: Reporting about internally sensitive news, by news outlets' types.



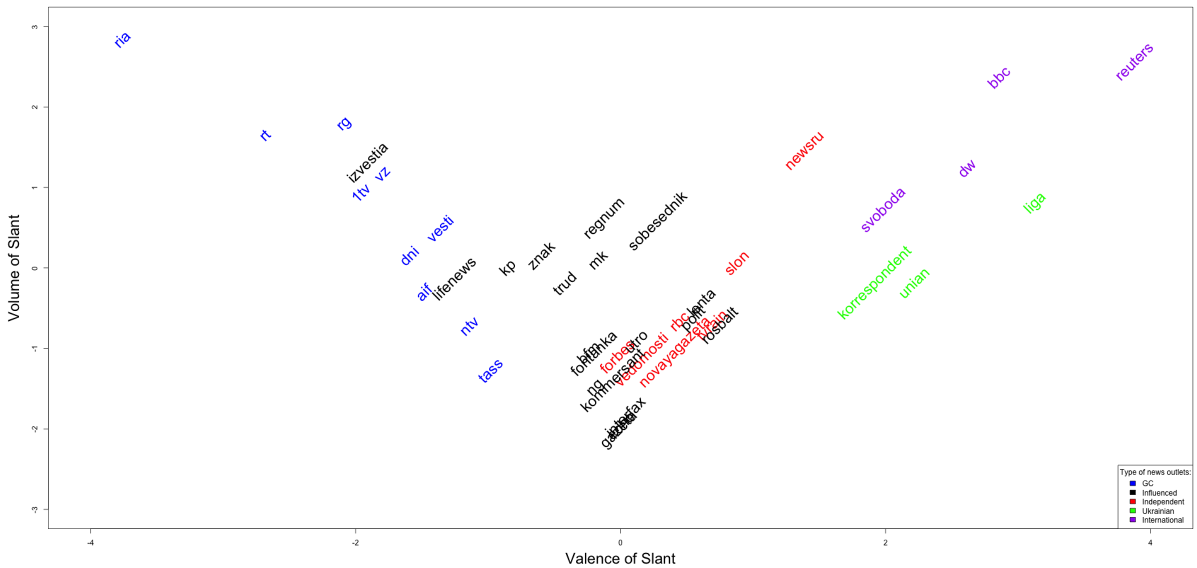
Each text string represents a position of a news outlet. We remove five news outlets for which we have only information about the titles and about the text of the articles.

Figure 13: Reporting about the Ukraine-crisis news by news outlet type.



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

Figure 14: Ideological positions of the news outlets in the Ukraine-crisis news coverage.



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

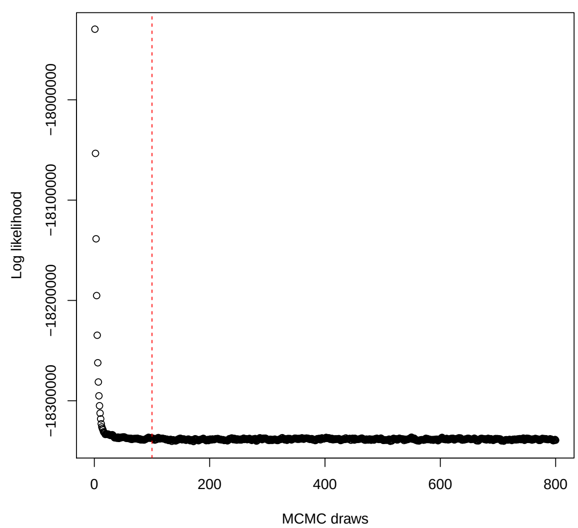
9.8 MCMC Estimation

We estimate the demand parameters by simulating from the posterior distribution defined in Section 5.3 and use the choice data from a set of consumers defined in Section 5.2. Given the large amount of choices that consumers make and a substantial number of choice alternatives (37), the estimation process requires substantial time and RAM memory resources. We use the hybrid MCMC sampler *rhierMnlRwMixture* from the 3.1 version of the *bayesm* package in R (Rossi et al., 2005). The sampler in this package is written in Rcpp, an R library that allows the integration of R and C++ languages, which substantially speeds up the computational process. Still, given the number of consumer choices and alternatives, estimation takes a significant amount of time. We run the MCMC procedure for the full sample of consumers for 20,000 iterations, storing every twenty-fifth observation for the memory reasons, leading to 800 saved MCMC draws. Estimation is complete in 12 days and requires 800 Gb of RAM. Figure 15 shows the log likelihood of the MCMC draws. We throw away the first 100 saved draws (2,500 actual draws) to remove the effect of the starting point on the sampler. Figure 16 shows an overview of the evolution of the $E(\theta)$ draws.[96] We treat the last 700

[96]We omit the draws of π_i since it simply ensures that consumers never revisit the news outlet on the same day.

MCMC draws as draws from the stationary distribution.

Figure 15: Log likelihood of the 800 kept MCMC draws.



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

Figure 16: MCMC draws of $E(\theta)$.