



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2016

---

## Censorship of the Arab Spring in MENA media

Ciocan, Dumitru ; Wüest, Bruno

**Abstract:** The press should ideally be the eyes, ears and voice of the public in any state. However, freedom of the press varies across states and is especially lacking in autocratic ones. This paper asks how the press in autocracies tackles the challenge of reporting on contentious mobilization, i.e. protests events that threaten the very survival of the regime. For this, it relies on count and structural topic models applied to an original dataset of roughly half a million newspaper articles published before and after the events of the Arab Spring (January 2009 – December 2011), and on new protest event data from the Mass Mobilizations in Autocracies Dataset. We find that both the extent of coverage and its content is influenced by the overall degree of censorship in MENA countries. Moreover, threats to authoritarian regimes, measured both as intensity of domestic protests and intensity of protests across the wider MENA region, also influence the coverage of the issue. We also find that these effects are stronger for state owned newspapers.

Posted at the Zurich Open Repository and Archive, University of Zurich  
ZORA URL: <https://doi.org/10.5167/uzh-143872>  
Conference or Workshop Item

Originally published at:

Ciocan, Dumitru; Wüest, Bruno (2016). Censorship of the Arab Spring in MENA media. In: European Political Science Association. Annual Meeting, Brussels, 23 June 2016 - 25 June 2016.

# CENSORSHIP OF THE ARAB SPRING IN MENA MEDIA

**DUMITRU CIOCAN**  
University of Zurich  
[ciocan@ipz.uzh.ch](mailto:ciocan@ipz.uzh.ch)

**BRUNO WUEEST**  
University of Zurich  
[wueest@ipz.uzh.ch](mailto:wueest@ipz.uzh.ch)

Paper prepared for presentation at the 6<sup>th</sup> Annual General Conference of the  
European Political Science Association  
23-25 June 2016  
Brussels

## **Abstract**

The press should ideally be the eyes, ears and voice of the public in any state. However, freedom of the press varies across states and is especially lacking in autocratic ones. This paper asks how the press in autocracies tackles the challenge of reporting on contentious mobilization, i.e. protests events that threaten the very survival of the regime. For this, it relies on count and structural topic models applied to an original dataset of roughly half a million newspaper articles published before and after the events of the Arab Spring (January 2009 – December 2011), and on new protest event data from the Mass Mobilizations in Autocracies Dataset. We find that both the extent of coverage and its content is influenced by the overall degree of censorship in MENA countries. Moreover, threats to authoritarian regimes, measured both as intensity of domestic protests and intensity of protests across the wider MENA region, also influence the coverage of the issue. We also find that these effects are stronger for state owned newspapers.

## **Keywords**

Arab Spring, Censorship, Contentious mobilization, MENA, Topic models

## Introduction

Information on political protests can significantly lower the revolutionary thresholds through bandwagon processes and informational cascades (e.g. Granovetter, 1978; Kuran, 1991; Lohmann, 1993). More precisely, information on protests gives cues to the broader population about the strength or weakness of the regime and can make individuals more likely to join protests. This, in turn, increases the incentives for autocratic regimes to suppress any reports on political protests. In a world of the Internet and mobile phones, however, an autocratic regime can hardly suppress all information on contentious mobilization, which is why it increasingly needs to manage information on protests. This means to influence the bias of reports on these events according to its own interests. This paper seeks to explore the variance in reporting practices of the media in autocratic regimes of the Middle East and North Africa (MENA) and answer the question how they tackle the challenge of reporting on events that threaten the very survival of the regime.

For this, it relies on machine learning to extract texts on contentious politics from an original dataset of roughly half a million newspaper paragraphs published by English-language newspapers in MENA countries before and after the onset of the Arab Spring (January 2009 – December 2011). Subsequently, count and structural topic models (Roberts et al. 2014) are applied to this original corpus in order to explore the expected reporting bias. Independent variables measuring the threat to the survival of the regime are derived from the Mass Mobilizations in Autocracies Dataset (Weidmann and Rød 2014). Variables measuring specific censorship events are derived from a novel dataset created especially for this study.

In this paper we establish that the MENA press does report on the issue of contentious mobilizations irrespective of ownership or overall press freedom, though both aspects influence the quantitative output on the subject. We also establish that the state-owned press reacts strategically to threats to the authoritarian regime it belongs to: when protests in the MENA region increase, the state owned press reduces coverage of the issue, but when domestic protests intensify it increases coverage of the issue. Furthermore, censoring specific news reduces the coverage of only the state owned newspapers.

Looking at the content aspects of reporting, we identify fourteen distinct topics employed by the MENA press when reporting on the issue of contentious mobilizations,

some of these topics being specific to the Arab Spring events. The prevalence of most of these topics varies significantly as protests (foreign or domestic) intensify and also depends on the degree of press freedom the media enjoys. The prevalence of these topics also varies significantly as censorship of specific news intensifies, but for some topics the effect is contrary to what was intended. A prime example for the bias present in the MENA press coverage of contentious politics is the ‘revolution’ topic, the prevalence of which drops the stronger a state’s control over the press is and increases the more intense protests (domestic or foreign) become. However, the prevalence of this topic rises significantly as a state intensifies its efforts to suppress specific domestic news (such as news about domestic protests), despite controlling for both the intensity of protests abroad and at home.

### **Media and contentious mobilization in authoritarian regimes**

Overthrowing an authoritarian regime is tremendously difficult, since it requires ordinary citizens to overcome paramount collective action problems. Among these problems figure the prevalence of preference falsification (Kuran, 1991), the presence of state repression and the absence of independent sources of information on the strength of both the regime as well as anti-regime sentiments among the population (Besley and Prat, 2006). In an authoritarian regime, the government therefore often colludes with the mass media to ensure that contentious mobilization is not reported. In other words, the fundamental function of the media to disseminate information can develop into a perilous risk for a regime when it comes to the coverage of political protests. Ultimately, information on anti-government protests give signals about the weakening of the regime. If the government is not able to censor reports on protests to a high degree, people update their beliefs about the weakness of the regime and, in turn, are more likely to join protests (Bhattacharyya and Hodler, 2015).

Formal models and empirical assessments of contentious mobilization suggest that ‘informational cascades’ play a crucial role in overcoming collective action problems (e.g. Lohmann, 1993). This literature highlights how information about the aggregate intensity of contentious mobilization conveys informational cues about the citizen’s political preferences, and how these cues trigger changes in the beliefs about the value of participating in anti-regime protests. The larger an existing opposition movement, the larger are the incentives for participation, since the likelihood of

repression for siding with the opposition is decreasing. More specifically, there is a ‘revolutionary threshold’, which denotes the point at which the costs of supporting an actually despised regime outweigh the costs of joining the opposition movement. Moreover, a revolutionary cascade or ‘bandwagon’ ensures that others join the opposition as well, and “authoritarian regimes that once appeared unshakeable can see their support crumble in no time” (Crabtree et al., 2015). This is why it is so important for regimes to maintain control over the production and dissemination of information.

Citizens generally lack the opportunity to monitor contentious mobilization directly, so they have to rely on information from the media to update them on political activities (Bairrett, 2015). Hence, protests in autocracies occur only when those opposed to the regime perceive that dissatisfaction is wide-spread and government repression ineffective (Kim, Whitten-Woodring and James, 2015). In a world of the Internet and mobile phones, many regimes can hardly suppress all information on collective action, which would be the optimal solution for any authoritarian regime (but see King et al. 2013 on China). During the Arab Spring, most MENA countries lacked free media. Hence, professional journalists, bloggers, and ordinary citizens who criticized the authorities were censored, fined, imprisoned, harassed, physically threatened or hurt, or worse (Kim et al. 2014). Nevertheless, many journalists and citizens persisted in spreading information about contentious mobilization, especially via social media (Gohdes, 2015). Nowadays, such regimes are thus rather forced to manage information on protests. So it seems plausible that a regime would want to induce a bias in the reporting on political protests to tend to their own interests (Entman, 2004:5). Empirical evidence also substantiates this claim, as demonstrated by Baum and Zhukov (2015) who find that the media in non-democracies underreport peaceful protests by regime opponents, overreport violent ones and largely ignore government atrocities. Thus, if authoritarian regimes were willing to let the people be informed about political protests at all, there are several strategies that can lower the risk of revolutionary cascades. The most straightforward option for the regime is the portrayal of protests as illegitimate (e.g. 'terrorism', 'crime'), insignificant (e.g. 'isolated event') or easy to suppress. The framing of protests in such a way, especially by the state-owned media, has been documented for example in the case of the Egyptian press (Hamdy and Gomaa, 2012). A more subtle strategy is to let media report about protests without interference for as long as they don't portray them as revolutionary or in a positive light. The practice of such self-censorship during the Arab Spring has been documented even for MENA

countries enjoying a relatively free press (Alkazemi and Wanta, 2015). In general, it can be assumed that the induced bias in the reporting on contentious politics is of a nature, which lowers the likelihood of individuals to join in mass mobilizations against the regime (Gehlbach and Sonin, 2014).

Given the considerations above, we first pursue the question of whether the media in MENA countries is indeed biased with regards to coverage intensity. More precisely, we ask ourselves whether the amount of coverage contentious mobilizations receives in the media is dependent on press freedom or threats to the survival of the authoritarian regime. Second, we pursue the same question with regards to the contents of the reports.

## **Methodology**

### **Data sources and preprocessing**

The time frame we examine begins in January 2009, which is roughly two years before the first protests of the Arab Spring in Tunisia and encompasses the Iranian Green Movement as a counterpoint to Arab Spring protests. The time frame ends in December 2011. To analyze how the MENA media covers contentious politics, and to gain a handle on the censorship policies of MENA states, we rely on three new datasets, as well as the Freedom House dataset.

The first dataset is comprised of articles published in the newspapers listed in Table A.1. Currently we have processed articles from thirty-three English-language newspapers from eleven MENA countries. We retrieved most newspaper articles from the databases LexisNexis and Factiva using simple but broad keyword searches<sup>1</sup>. Web-scraping the homepages of newspapers not available in the aforementioned databases yielded additional articles for our analysis. We then split the articles into paragraphs of similar length (ca. 150 words), which yielded a corpus of 523'434 paragraphs. A manual evaluation of a random sample of 1'334 documents revealed that 31.2% of the texts covered contentious politics. Such relevant texts are those that provide coverage of public, politically motivated and unconventional actions (Dolezal et al. 2012). This definition thus excludes for example institutional processes (e.g. elections or court

---

<sup>1</sup> The keyword string for the different newspaper database was an adaptation of “protest! OR demonstrat! OR manifest! OR marche! OR marchi! OR tumult! OR parade OR rall! OR picket! OR (human chain) OR riot! OR affray OR unrest! OR clash! OR vigil OR strike! OR boycott! OR block! OR sit-in OR squat! OR mutin! OR revolt! OR revolution! OR oust! OR rebel! OR dissent! OR turmoil! OR violen!” depending on the options available for Boolean operators and truncation wildcards.

cases) and most protests by elites (e.g. diplomatic protests). At the same time the definition includes actions such as strikes, collection of signatures, petitions, demonstrations, marches, protest camps/meetings/vigils/rallies, boycotts, cyber-attacks, hunger strikes, refusals of payment, blockades, squatting, symbolic violence against objects or persons, bomb and arson attacks, sabotage, riots, destruction of private or public buildings, violence against persons and clashes with the police. Since we are interested in how newspapers cover political protest events, we do not evaluate whether the events described actually took place or not. Consequently, we also include threats to mobilize and planned events.

In a pre-processing step, we converted all words of the 523'434 documents to lowercase, applied Porter's stemming algorithm and removed punctuation, stop-words and sparse words. Furthermore, we substituted abbreviations and *n*-words relating to countries, organizations and political parties with corresponding single tokens (e.g. "UAE" to "UnitedArabEmirates). Using the manually coded paragraphs as the training set, we classified all paragraphs of our corpus as either relevant or irrelevant by using an ensemble consisting of five support vector machine (SVM) classifiers. An ensemble is a tool that aggregates the results produced by multiple classifiers, which may use different classification algorithms (Dietterich, 2000; Collingwood and Wilkerson, 2012). The text classification was implemented using the R package *classyfire*, which has several advantages compared to prior approaches, ranging from increased computational speed, to automated bootstrapped training, automated parameter optimization and permutation testing (Chatzimichali and Bessant, 2016). The SVM ensemble achieved an F-Score of 0.84 (0.85 precision and 0.83 recall) when tested against the manually coded validation set. Using this ensemble we obtained a corpus of 104'183 documents dealing with contentious politics. We also collected ownership information for the newspapers included in our corpus. We relied primarily on WAN-IFRA country reports (World Association of Newspapers, 2010), but crosschecked and complemented the data with information provided by the homepages of the respective newspapers and other sources. We differentiate between outright state-owned newspapers and other (semi-) independent ones.

The second dataset, the protest event dataset, stems from the collaboration in the "Mass Mobilization in Autocracies Dataset" project (Weidmann and Rød, 2014). A

first release of the dataset is expected later this year<sup>2</sup>. This dataset includes the date, location and side (pro- or anti-government) of all protest events for the MENA region for the time period of interest, as reported by *Agence France Presse*, *The Associated Press* and *BBC Monitoring*. The independent variables *domestic protests* and *MENA protests* operationalize the concept of threat to the survival of the authoritarian regime. They count the number of events taking place each week in a given country and respectively across the entire MENA region.

The third dataset comprises censorship events against press, radio and TV. The raw data stems primarily from *Reporters without Borders* news releases and is complemented by *U.S. Department of State* and *Freedom House* country reports, as well as several Human Rights organizations mentioned in these news and reports. Censorship events include arrests, attacks, blocking of information dissemination, confiscating or destroying equipment, fines, gag orders, kidnappings, killings, laws, lawsuits, unemployment, temporary or permanent shut-downs, threats and sexual harassment. The aim of such actions ranges from the national over the organizational and down to the individual level. Coding events at the national and organizational level is straightforward, but events at the individual level are often less clear. We decided to code an event at this level whenever a journalist is hindered from collecting or publishing some news data. We also differentiate between *violent* and *not violent* events, violent events being any actions that include physical violence against journalists and even the rather marginal cases where journalists were arrested or otherwise detained against their will. Finally, we also differentiate between acts of *retaliation* against journalists or organizations, provoked by some news reporting in the past, and acts of *suppression*, which do not target specific journalists or organization, but aim to hinder the (timely) reporting on some current event.

Due to the very limited number of censorship events at the national level, we exclude this particular variable from our analysis. All other variables have been recoded as count data, where we count the number of relevant events which took place during the month prior to the week of publication of an article in the respective country. In addition to the aforementioned censorship variables, we also constructed a general *censorships* variable counting all events in a similar manner. We also take into consideration the overall level of censorship in a country during a given year by

---

<sup>2</sup> Article manuscript, codebook and eventually the dataset are available at <http://www.cnc.uni-konstanz.de/research/mmad/>



including the three Freedom House indices measures *economic*, *political* and *legal* environments in our analysis.

### Estimations

The question on what influences the amount of ink the MENA press dedicates to contentious politics is explored by negative binomial regressions. We estimate a total of 8 models, where the dependent variable is the number of paragraphs on contentious politics published during a week by a given newspaper (see Table A.2). The dependent variable is characterized by a high number of weeks with no such paragraphs. A high count of 0 in the dependent variable usually leads to an over-dispersed estimation if the usual Poisson model were to be used (Long and Freese, 2006). As the highly significant log-likelihood ratio test for Model 1.0 in Table A.2 indicates, a negative binomial regression is therefore recommended in this particular case.

The basic model (Model 1.0) includes the independent variables *economic*, *political* and *legal control*, *ownership*, *domestic protests*, *MENA protests* and the *censorships* variable. Models 2.0, 3.0 and 4.0 expand on the basic model by substituting the *censorships* variable with the variables detailing the different kinds of censorship events, namely *suppression* and *retaliation* in Model 2.0, *violence* and *non-violence* in Model 3.0 and *individual* and *organization* in Model 4.0. These models are in turn expanded in Models 1.1, 2.1, 3.1 and 4.1 through the inclusion of interactions between *ownership* on the one hand, and the censorship and protest variables on the other. Like the number of paragraphs, all variables counting protests and censorship events are highly skewed towards 0, which is why we log-transformed these variables in our models using the formula  $\log(x+1)$ . All models include as controls country fixed effects and a weekly time trend to minimize serial correlation.

The question of how the MENA press writes about contentious politics and what influences this decision, we explore with a structural topic model (STM) (Roberts et al., 2014), which allows us to estimate document probabilities for latent variables, called *topics*. STM builds on the Latent Dirichlet Allocation, a hierarchical mixed-membership model in which the document-topic and word-topic probabilities have a common prior drawn from a Dirichlet distribution (Blei, Ng and Jordan, 2003). One of the STM's major innovations is that the prior distribution of topics (i.e. topic *prevalence*) can be influenced by covariates. The STM analysis includes the same covariates as the

negative binominal regression.

A crucial decision in every application of a topic model pertains to the granularity, i.e. the number of topics, which needs to be determined a-priori. A topic model with too few topics will produce overly broad, diffuse topics, while a model with too many topics will result in many small, hardly distinguishable topics (Greene and Cross, 2015). An increasingly popular strategy to resolve this problem is to compare the coherence of different topic models. To this purpose, we use *word2vec* (Mikolov et al., 2013), which learns and aggregates term similarities through a shallow neural network process. Word2vec suggests a granularity of 14 for our corpus and a candidate range of 13 to 16 topics (see Figure A.1), these numbers resulting from comparing the coherence within and between the vectors of most probable words for each topic model.

## **Results**

### **Media attention to contentious mobilization**

A first overview of media coverage on contentious politics clearly shows the importance of the Arab Spring (see Figure 1). Already before the first turning point, namely the ousting of Tunisian ruler Zine El Abidine Ben Ali, the number of weekly reports on contentious mobilization starts to climb and reaches its peak with the onset of the NATO intervention in Libya. The regimes in the MENA region were thus largely incapable (or unwilling) to suppress most of the information on political protests even in traditional media, which are relatively easy to censor, especially in the frequent case of state ownership. This result sharply contrasts the sophisticated and highly effective efforts to suppress information on collective action on all channels including the Internet in countries such as Iran or China (King et al. 2013).

The results of the negative binomial regressions (see Table A.2.) show that according to the Aikake information criterion regressions modeling interactions between variables are marginally better than their respective counterparts. However, irrespective of which model we look at, the two Freedom House indicators measuring the state's economic and legal control over the media are highly significant, while the third indicator, political control, is also significant in all but the last model.

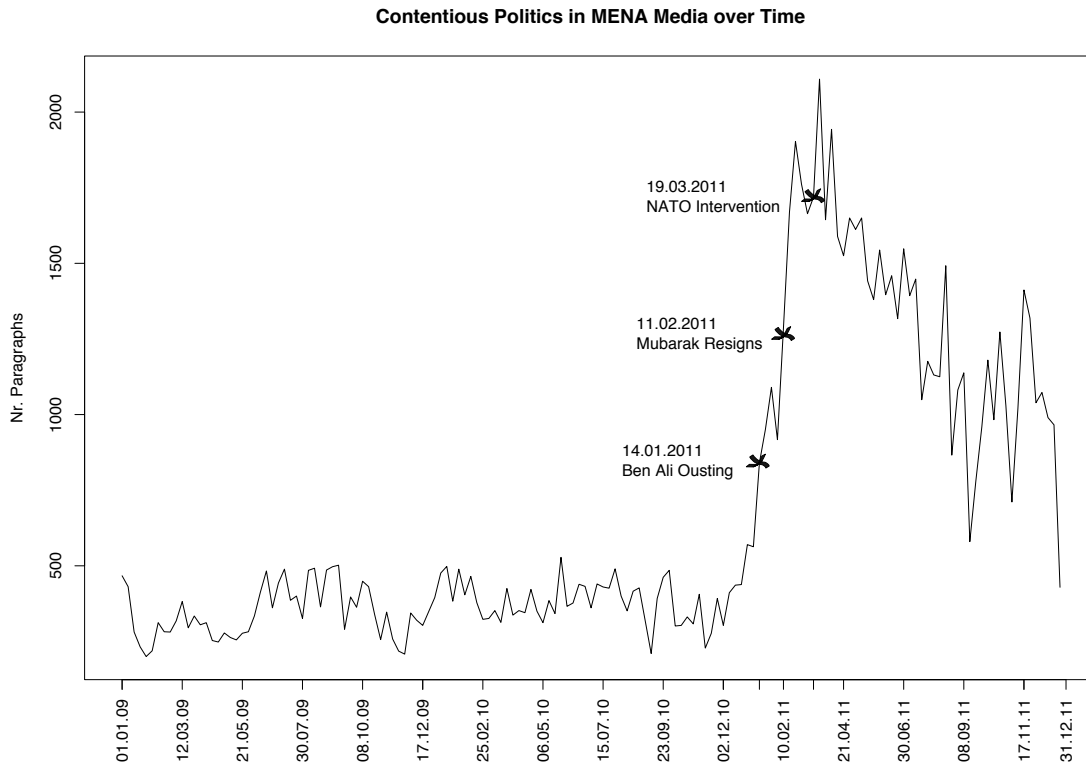


Figure 1: *Coverage of contentious politics in MENA media over time*

As shown in Figure 2, which plots the effects of these variables according to Model 1.1, these variables influence our dependent variable in quite different ways. Economic control has by far the greatest influence and its negative sign conforms to our expectation that tighter economic control over the media diminishes the number of paragraphs published concerning contentious politics. Contrary to our initial expectations, tighter political and legal control does not lead to less publishing, but actually increases the number of paragraphs on contentious politics.

The domestic protest variable is similarly highly significant across the board, having a positive effect on the dependent variable. The main effect of this variable remains fairly constant even when we consider its interaction with the newspaper ownership variable. Furthermore, as Models 1.1 and 4.1 indicate, state owned newspapers react significantly different than other newspapers, namely printing less paragraphs on contentious politics, when domestic protests are absent or only few in number (see Figure 3).

Contentious Politics and Press Freedom (Model 1.1)

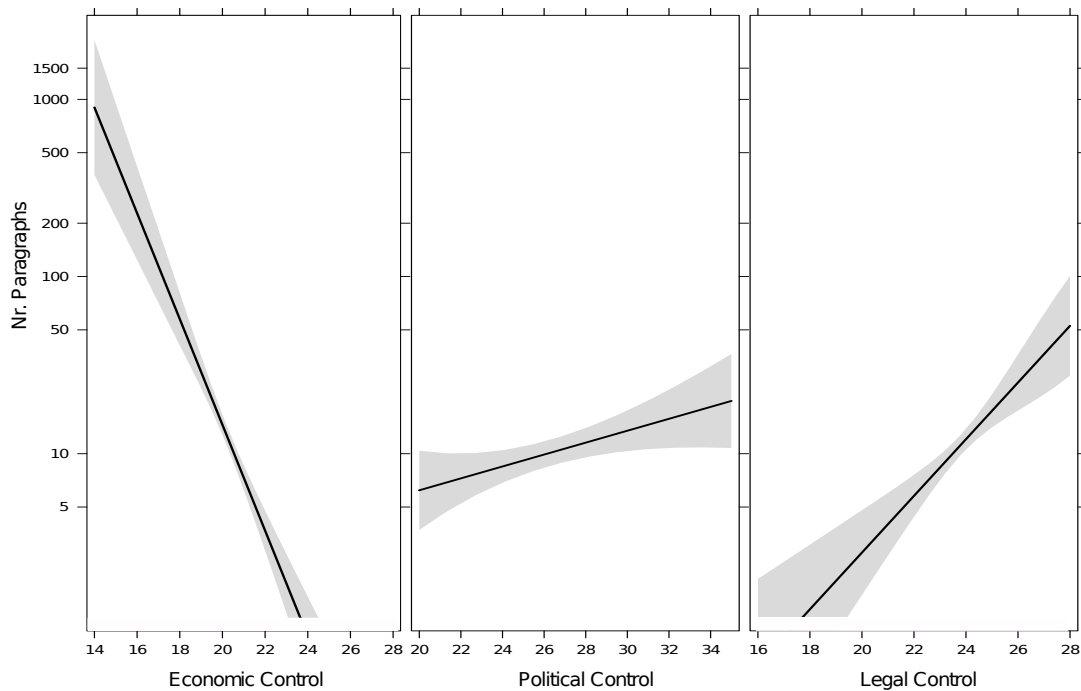


Figure 2: Relationship between overall political, economic and legal control over the media and the newspapers' coverage of contentious politics.

The MENA protests variable has a negative effect on the dependent variable in each model and the effect is significant in each model lacking interactions. In models that do consider interaction effects, the main effect of the variable falls below the statistical threshold of significance, but the interaction with the ownership variable shows that state owned newspapers write significantly less about contentious politics as protests in the MENA region intensify (see Figure 3).

Turning to the censorship variables we find more complex relationships. Looking at the results of Models 1.0 we find that the general censorship variable has a slight negative effect on the dependent variable, but this effect doesn't reach the statistical significance level. However, when we interact this variable with the ownership variable both the main effect and the interaction effect become significant. As evidenced in Figure 3, there is little difference between state owned and other newspapers when there are no censorship events of any kind, but as such events increase in number state owned newspapers react by publishing less paragraphs on contentious politics, while other newspaper publish more. Differentiating between censorship events motivated by retaliation and suppression, as in Models 2.0 and 2.1, leads to mixed results. In Model 2.0 only the retaliation variable has a slight, but

significant effect, while in Model 2.1 the main effect of only the suppression variable is positive and significant. Furthermore, the interactions of both variables with the ownership variable are statistically significant, state owned newspapers publishing fewer paragraphs as censorship events increase than other newspapers. Differentiating between violent and non-violent censorship events we find that the violent censorship variable has no statistically significant impact on the dependent variable. The non-violent censorship variable, however, has a negative and statistically significant effect on the dependent variable in Model 3.0. Model 3.1 shows that this effect is due to the interaction with the ownership variable, state owned newspapers reacting by publishing less as such events intensify. Lastly, when differentiating between censorship events against individuals and organizations we find that only the interaction of the individuals variable with the ownership variable is statistically significant, state owned newspapers publishing, as in the other models, less as these events intensify. Figure A.2 plots the interaction between newspaper ownership and the different censorship variables.

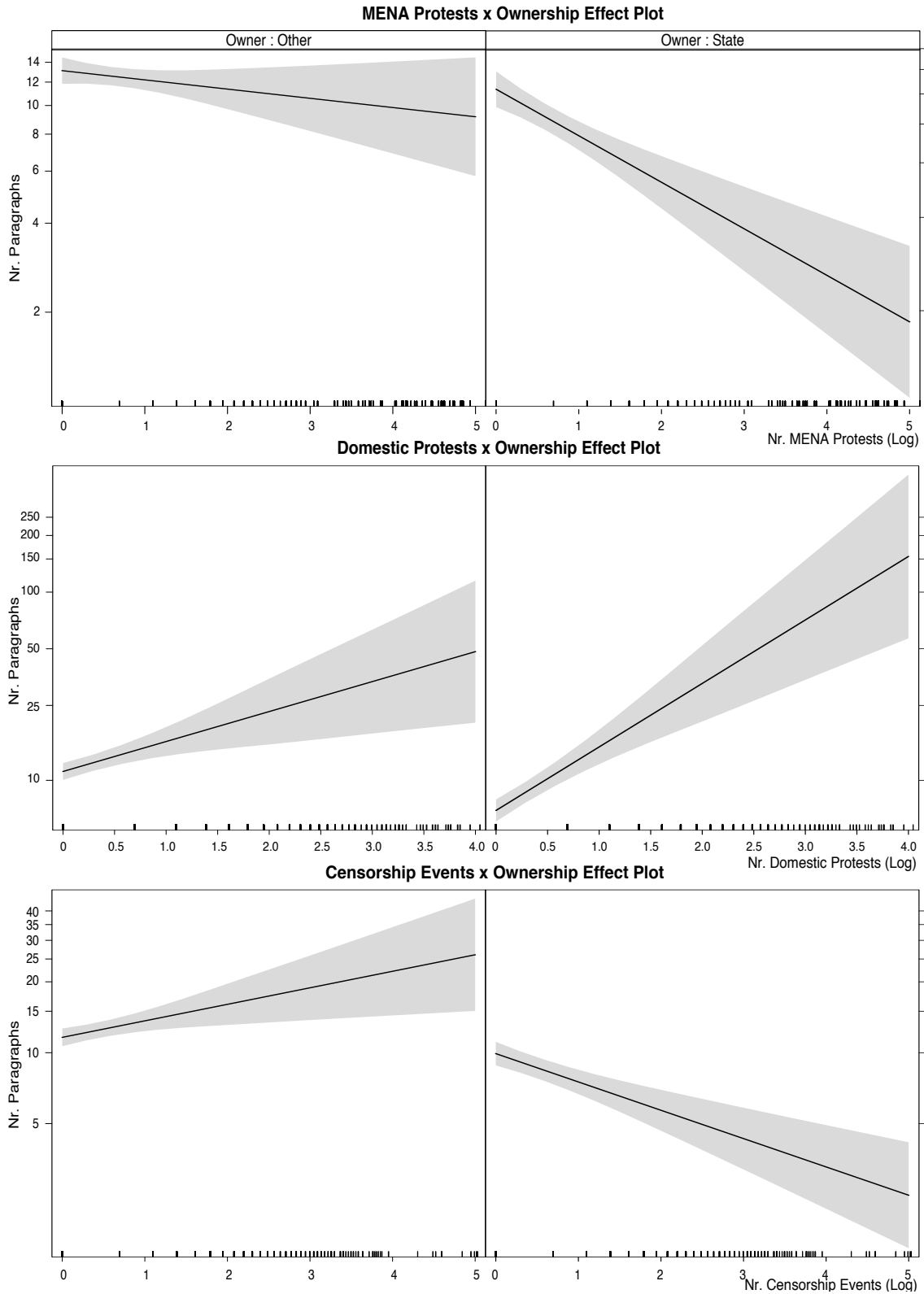


Figure 3: *Reporting bias of state-owned newspapers as threats to the regime and threats to journalists increase*

## Media bias with regards to contentious mobilization

Thus far, we have determined that differing characteristics of autocracies do influence the amount of coverage that contentious politics receive in the press to differing degrees and that direct ownership of newspapers is a highly effective means to impose compliance with the regime's interests. At the same time, while effective at censoring news from beyond their borders, they are less so when contentious politics intensify on the home front. This leads to the question of whether the content of the news is also subject to similar influences.

The remaining analyses will be concerned with this follow-up question. Table 1 shows the results of a fourteen-topic model in terms of a list of the most probable word stems. This list was used to ex-post annotate the labels shown in the second column.

The topics identified fall in three categories, the first consisting of broad topics related to the Arab Spring (i.e. *peaceful protests*, *violent protests*, *democracy*, *revolution*, and *political implications*), the second consisting of country specific topics (i.e. *Iraq*, *Libya*, *Syria* and *Yemen*) and the third being other broad topics, but with little bearing on the Arab Spring (i.e. *judicial process*, *economy*, *regional conflicts* and *civilian perception*). Topic nr. 2 in our model is characterized by words relating to protests (e.g. *protest*, *demonstr*, *march*, *ralli*, *movement*, *gather*), the motive of the protests (e.g. *call*, *govern*, *demand*, *anti*, *reform*, *pro*, *support*, *right*), the participants in the protests (e.g. *activist*, *poeopl*, *opposit*, *group*), the timing and location of the protests (e.g. *street*, *friday*, *week*, *squar*). We label this topic *peaceful protest* in order to differentiate it from topic nr. 5, *violent protest*, which associates protests with violence (e.g. *kill*, *fire*, *clash*, *dead*, *injur*, *wound*) and the mobilization of state's resources to meet this violence (e.g. *polic*, *forc*, *secur*, *armi*, *troop*). Topic nr. 3 is defined by words relating to wielders of political power (e.g. *presid*, *leader*, *primeminist*, *rule*, *power*, *govern*), but also institutions associated with democracy (e.g. *elect*, *parti*, *opposite*, *constitut*, *will*, *vote*, *parliament*), the jasmine revolution (e.g. *ben*, *ali*, *tunisia*) and the South Sudan independence referendum (e.g. *sudan*). Given the broad spectrum of this topic and the prominence of elections in it, we label it as *democracy*. Topic nr. 6 also touches upon events in two different countries, being characterized by words we strongly associate with the Egyptian revolution (e.g. *revolut*, *protest*, *egypt*, *cairo*, *tahrir*, *squar*, *hosni*, *mubarak*), but also with the Iranian green revolution (e.g. *iranian*, *revolut*, *protest*), which we thus label as the *revolution* topic. In topic nr. 7 the revolution aspect of contentious politics is also present (e.g. *revolut*), but there is a time

component (e.g. *now, time*) and an action perspective (e.g. *will, can, need, must, take, chang, support*) embedded in a broader political context (e.g. *world, state, countri, region, arab*). We label this topic the *political implications* of the Arab Spring.

The topics nr. 4, 9, 10 and 11 we labeled as *Iraq, Libya, Yemen* and respectively *Syria*, since these topics abound of country-specific references. The most prominent words of the *Iraq* topic paint the image of localized acts of terror (e.g. *kill, attack, bomb, suicid, explosion, blast*) with specific targets (e.g. *target, offici, polic, secur, soldier*), but not naming the perpetrators. The Arab Spring events figure strongly in the *Libya* topic, both domestic and foreign parties fighting the regime being highlighted (e.g. *rebel, nato, unitedn, benghazi*). We also find references to the most defining actions undertaken by these parties, such as the no-fly zone (e.g. *air*). The *Syria* topic is defined not as much by the Arab Spring, as by Syria's long lasting interests and involvement in Lebanon. We thus encounter next to the multiple references to Syria and its leader similar references to Lebanon and such describing the relationship between them (e.g. *specialtribunalforlebanon, leagu, hezbollah*). In the case of the Yemen topic, we also encounter references to domestic challengers to the rule of the long time autocratic leader from before the Arab Spring events (e.g. *houthi, tribal, rebel*), as well as foreign interests and involvements in local politics (e.g. *saudi, saudiarabia, gulfcooperationcouncil*).

Topics nr. 1, 12, 13 and 14 cover other aspects related to contentious politics. In topic nr 1, the *judicial process*, we encounter words that cover the apprehension of individuals (e.g. *arrest*), the accusations brought against these individuals (e.g. *accus, alleg, crime, charg*), their defense (e.g. *court, investig, lawyer, law, trial*) up to their sentencing (e.g. *sentenc, prison, releas*). In topic nr. 13, the top of the list is held by the United States, followed by names of prominent terrorist groups (e.g. *qaeda, taliban*), names of the countries harboring them (e.g. *afghanistan, pakistan*) and the explicit naming of terrorism, which is why we labeled this topic *war on terrorism*. Another highly coherent topic is topic nr. 14, labeled as the *economy* topic due to the fact that words such as strikes, unions, workers and employees figure high on the list of words defining it. We named topic nr. 14 the *civilian perception* of contentious politics since this topic links the average civilian (e.g. *women, peopl, young, famili, men, children, old*) to places one frequents (e.g. *street, school, home*) and to the media (e.g. *media, live, televis*), while potential dangers from contentious politics, or any mention of contentious politics for that matter, are not encountered in the most prominent words of



this topic. Lastly, topic nr. 8 deals with two other regional conflicts, their causes being unrelated to the Arab Spring, namely the Israeli-Palestinian and the Turkish-Kurdish conflicts.

Table 1: Label and list of characteristic words for the 14-topic model on paragraphs covering contentious mobilization

Topic	Label	25 most probable words (stems)
1	Judicial Process	arrest, court, right, prison, releas, charg, human, investig, accus, alleg, case, report, year, trial, state, law, also, lawyer, polic, crime, secur, sentenc, death, offic
2	Peaceful Protest	protest, demonstr, call, govern, demand, march, activist, peopl, ralli, bahrain, opposit, anti, reform, street, friday, thousand, support, movement, gather, group, pro, right, week, squar
3	Democracy	elect, parti, govern, presid, new, polit, leader, opposit, constitut, ali, will, group, primeminist, rule, year, sudan, power, vote, tunisia, ben, islamist, countri, parliament, nation
4	Iraq	kill, attack, bomb, peopl, two, offici, wound, polic, iraq, secur, suicid, three, soldier, explos, target, group, blast, provinc, one, four, violenc, civilian, car, citi
5	Violent Protest	polic, forc, citi, secur, peopl, kill, fire, clash, protest, resid, armi, violenc, town, two, least, dead, wit, report, area, injur, troop, one, wound, day
6	Revolution	egypt, mubarak, egyptian, cairo, militari, iran, presid, protest, hosni, tahrir, iranian, revolut, squar, former, council, offici, forc, say, rule, report, news, minist, state, armi
7	Political Implications	will, countri, arab, peopl, polit, regim, revolut, world, can, now, chang, support, state, power, one, time, like, region, mani, even, need, take, violenc, must
8	Regional Conflicts	israel, isra, palestinian, gaza, hama, border, turkey, turkish, strip, militari, kurdish, group, west, rocket, erdogan, fire, attack, bank, Jerusalem, kurdistanworkersparti, cross, fatah, war, offici
9	Libya	gaddafi, libya, rebel, libyan, forc, tripoli, nato, unitedn, leader, fight, militari, govern, civilian, air, benghazi, fighter, foreign, report, western, town, council, citi, muammar, told
10	Yemen	yemen, saleh, yemeni, presid, govern, saudi, abduallah, gulfcooperationcouncil, qaeda, ali, saudiarabia, sanaa, power, opposit, gulf, secur, rebel, countri, sana, houthi, fight, leader, office, tribal
11	Syria	syria, syrian, assad, lebanon, hariri, lebanes, hizbullah, damascus, arab, march, unitedn, specialtribunalforlebanon, govern, regim, presid, primeminist, beirut, say, bashar, leagu, hezbollah, report, polit, cabinet
12	Civilian Perception	one, year, women, peopl, show, old, mani, televis, young, famili, day, say, time, man, want, street, media, live, men, children, just, home, school, get
13	War on Terrorism	unitedst, pakistan, milit, qaeda, taliban, attack, afghanistan, offici, pakistani, militari, afghan, kill, forc, oper, govern, strike, terror, border, secur, bin, insurg, laden, troop, intellig
14	Economy	strike, oil, worker, million, union, countri, year, govern, compani, unrest, month, bank, day, work, employe, percent, aid, last, per, price, econom, week, trade, economi

As evidenced in Figure 4 the prevalence of the *democracy* topic does not depend on the level of economic or legal control over the media and the degree of political control has only a slight negative influence. In contrast, the prevalence of the *revolution* topic does depend on the degree of control over the media. Whether economic, political

or legal control, the difference in topic prevalence between the freest press and the least free one is about 10%, while in the least free ones the topic is basically non-existent. Figure 5 shows the prevalence of these two topics as a function of MENA protests, domestic protests and overall censorship events. On the one hand, we again find that the covariates have only a slight effect on the prevalence of the *democracy* topic, while on the other hand they exercise a strong effect on the prevalence of the *revolution* topic. The positive relationship between the *revolution* topic and the MENA protests or the domestic protests variables is to be expected given the news-worthiness of a (potential) revolution. Interestingly, the same positive relationship can be observed between the topic prevalence and the censorship variable. Not only is the topic more likely to be used the more censorship events take place in country, but the relationship is even stronger than for the other two variables. Thus, despite the fact that both topics address regime change, we find quite different relationships between the topic prevalence and the considered variables.

Another pair of related topics is formed by the *peaceful protest* and the *political implications* topics, both showing similar prevalence over time and, at least theoretically, one being the continuation of the other, contentious politics moving towards the sphere of normal politics. Looking at Figure 6, we see that the prevalence of the *political implications* topic doesn't depend on economic, nor on political control, but has a rather strong negative relationship with the legal control. However, unlike for the *revolution* topic, even in countries with the greatest legal control over the press this topic is estimated to make up almost 10% of the paragraphs on contentious politics. This topic is thus not subject to strong censorship. The prevalence of the *peaceful protests* topic, however, has a slight yet positive relationship with all three variables. Despite these differences, as Figure 7 shows, we find that the prevalence of both topics rises as MENA protests, domestic protests, as well as censorship events intensify. Again it is interesting to observe that as the number of censorship events grows, so does the prevalence of both topics.

Another example of paired topics is that of *Syria* and *Libya*. The *Syria* topic is subject to strong censorship similar to that of the *revolution* topic. For each variable of control over the press, we find that the prevalence of the topic diminishes as control grows. The difference in prevalence between the two extremes is about 10% and reaching in the one extreme 0%. Unlike the *Syria* topic, the *Libya* topic is subjected to less censorship. Economic control and legal control both exercise a strong negative

influence on the prevalence of this topic, while political control has only a negligible influence (see Figure 8). As evidenced in Figure 9, the *Syria* topic is the first case where the prevalence of the topic has a strong negative relationship with the three variables MENA protests, domestic protest and censorship events. These results suggest that when a country is facing high levels of domestic protests and steps up its censorship practices, this topic basically disappears from the press of the respective country. However, this is not the case for the *Libya* topic. The prevalence of the topic in a country's press rises when MENA protests or domestic protests intensify, but it sinks when censorship events increase.

There are also topics such as the *judicial process* and *Yemen* for which increased control over the press has a positive effect on the prevalence of the respective topic.

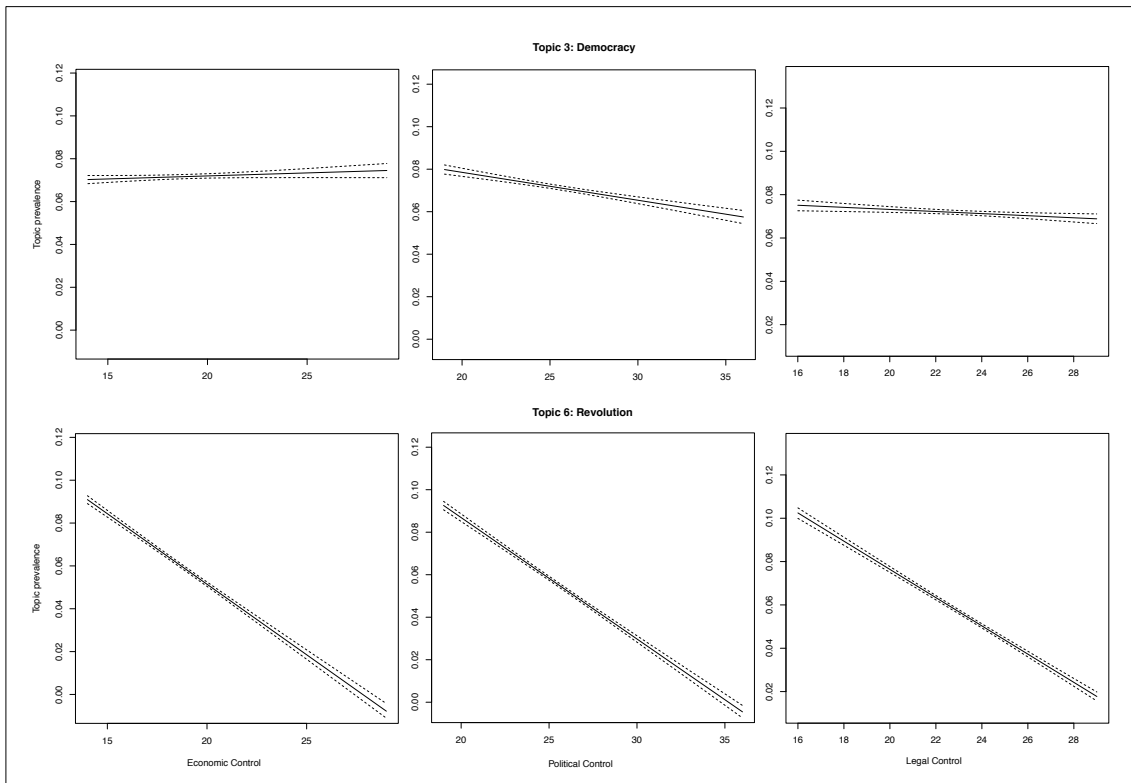


Figure 4: Topic Prevalence as a Function of Control over the Press (Democracy and Revolution)

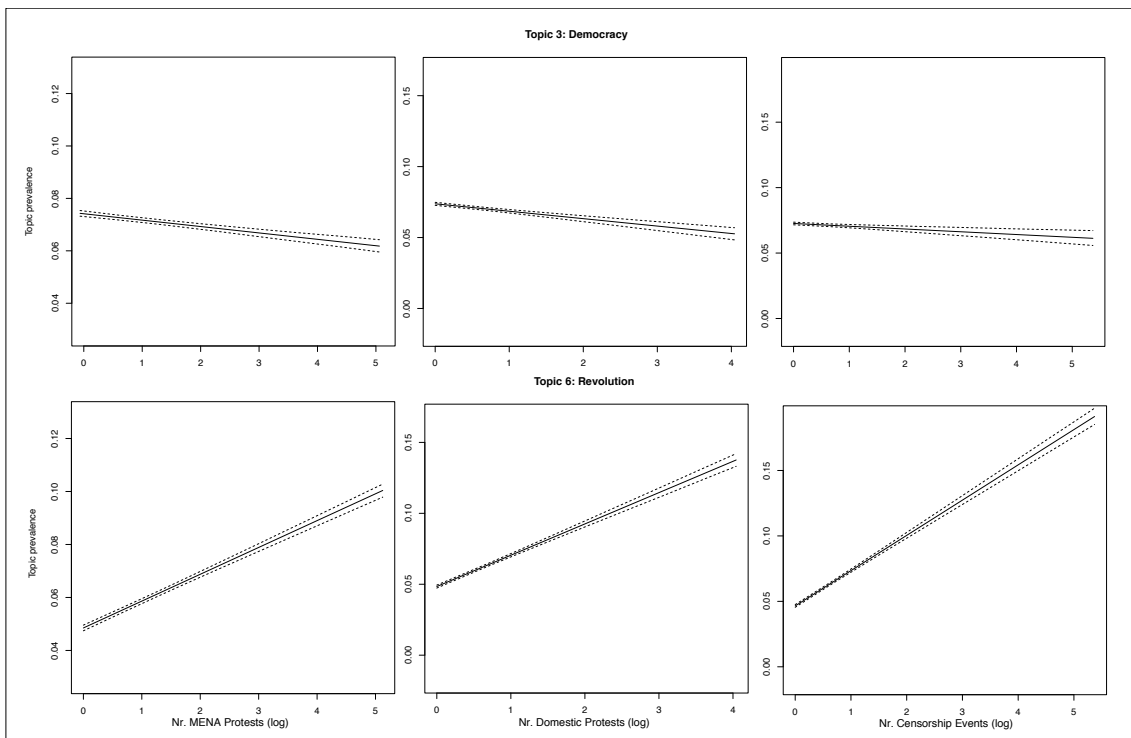


Figure 5: Topic Prevalence as a Function of MENA Protests, Domestic Protest and Censorship (Democracy and Revolution)

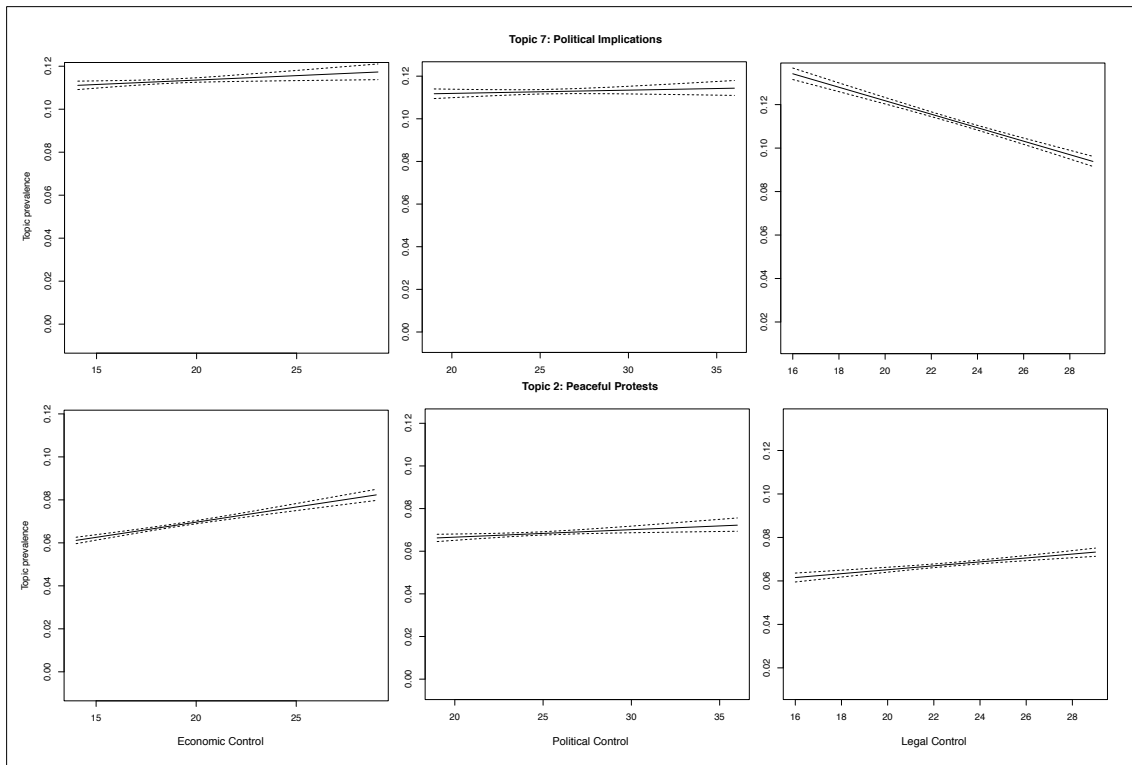


Figure 6: Topic Prevalence as a Function of Control over the Press (Political Implications and Peaceful Protests)

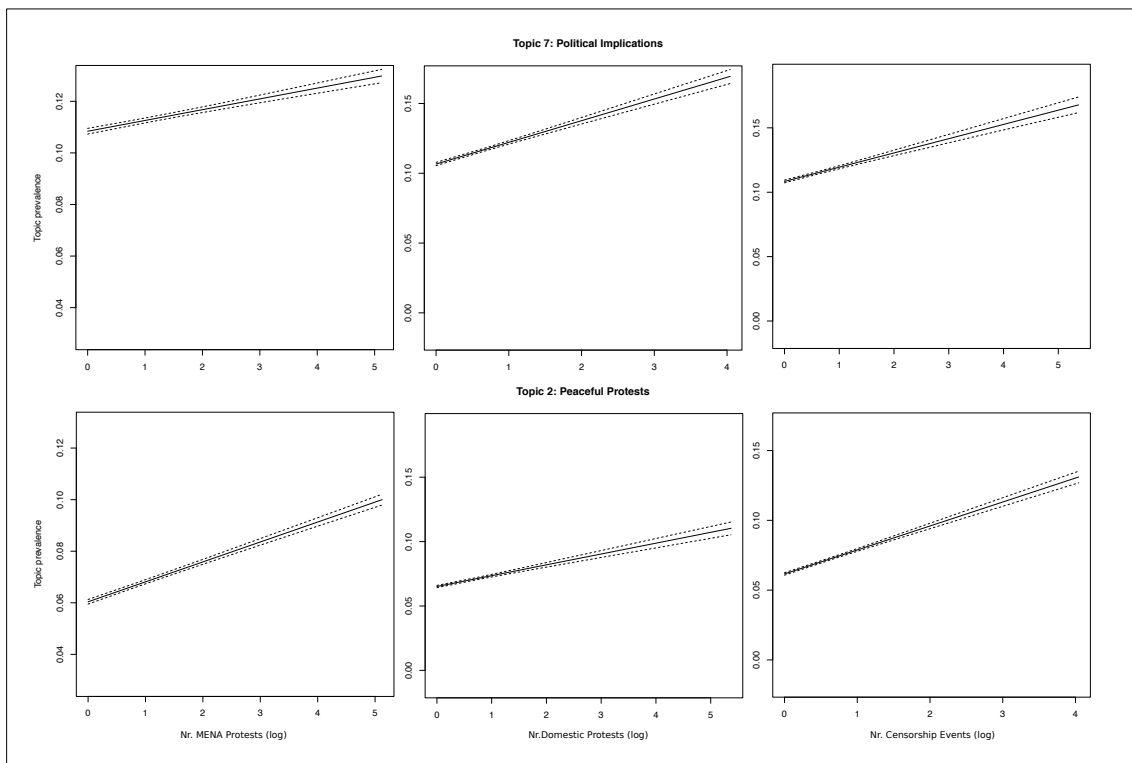


Figure 7: Topic Prevalence as a Function of MENA Protests, Domestic Protest and Censorship (Political Implications and Peaceful Protests)

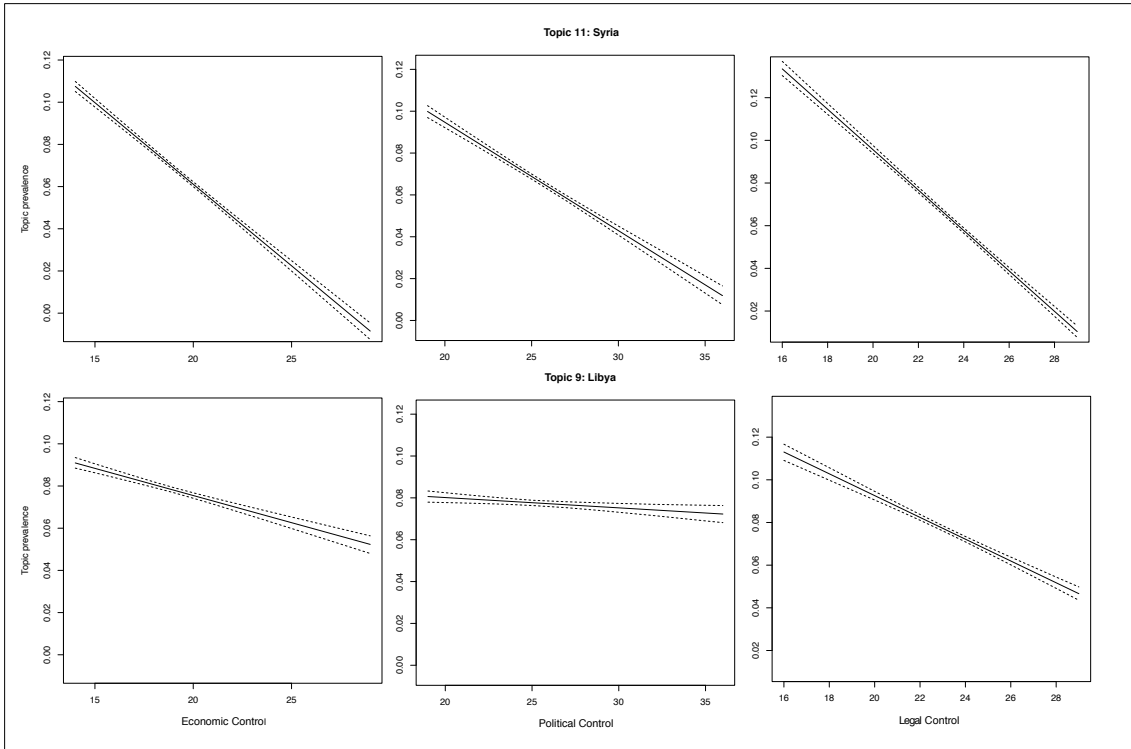


Figure 8: Topic Prevalence as a Function of Control over the Press (Syria and Libya)

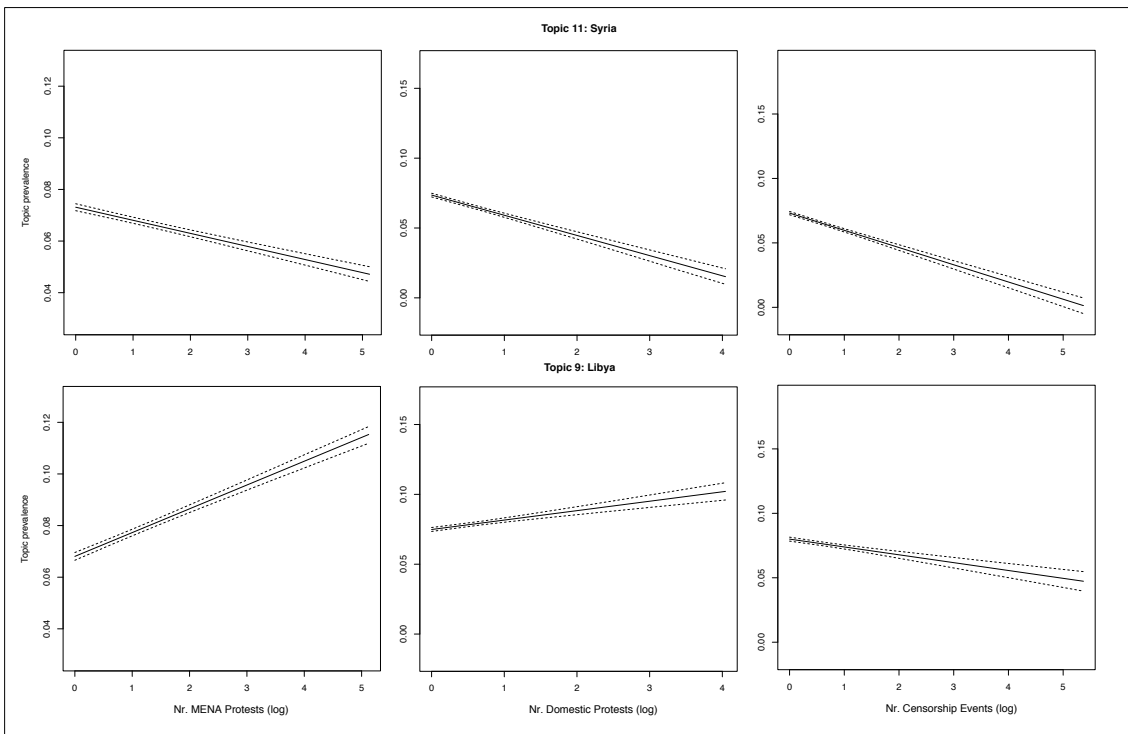


Figure 9: Topic Prevalence as a Function of MENA Protests, Domestic Protest and Censorship (Syria and Libya)

## Conclusion

Authoritarian regimes go to great lengths to limit what the media can write about and how to go about it. When faced with new and unexpected developments, one way to limit potential damage is to clamp down on the diffusion of the pertinent information. In the context of the Arab Spring uprisings this wasn't an option however. Information about the magnitude of the Tunisian protests, the success of the Jasmine Revolution, as well as the follow up events elsewhere, spread through multiple channels at nearly instantaneous speeds. Information spread through channels outside the immediate control of the states: social media, foreign TV channels and mobile phones. In this paper we ask ourselves how authoritarian regimes in the MENA region reacted in this novel situation on one channel they did have some degree of control over, namely the domestic press.

Our analysis provides a first step in better understanding what censorship policies they may have employed. Our original corpus of more than 100'000 paragraphs on contentious politics from English-language newspapers of the MENA region shows that the Arab Spring protests were indeed a much-discussed topic in the MENA press at the time. Furthermore, our data on censorship events shows that regimes reacted to these unexpected developments by actively blocking the reporting of news about domestic protests. We also establish that the most effective means that authoritarian regimes have at their disposal in order to limit the dissemination of unwanted information is the economic control over the media, namely a big market share of state-owned outlets, limiting access of the independent media to resources necessary for news production, use of bribery, allocating/withdrawing advertising and subsidies etc. However, greater legal and political control over the media has a positive effect on the extent of coverage of contentious politics by the press, though this effect is orders of magnitudes lesser than that of the economic control. This suggests that greater legal and political control over the press makes it easier or safer for journalists to report on contentious politics by spelling out the rules on how to do so, namely the hypothesized reporting bias. Furthermore, we find that domestic threats to the survival of the authoritarian regime force the media to report on the issue. Each of our models shows that when domestic protests intensify, the press writes significantly more about contentious politics. However, ownership of the newspaper is also an important factor. State-owned media is much more susceptible to the regime's wishes than other media. An increase in censorship events in a country has a deterrent effect only on the state-owned media. The

same media also stand out by actually reducing the coverage on contentious politics when protests in the region intensify. However, when domestic protests intensify the same media increases its coverage even more than that of other media, presumably to champion the state's preferred framing of the events. The count models also suggest that non-violent means of censorship and censorship aimed at individuals are highly efficient methods to ensure immediate compliance of the state owned media. One way to sum up these findings is that MENA states had little influence over the extent of coverage contentious politics received from the media not owned by them. However, the more autocratic a state is, the more media it seems to own.

Our STM identifies fourteen topics used by the press to report on contentious politics, including some of the topics we expected to encounter such as *revolution* and *democracy*, but not others such as an Arab Spring related *terrorism* topic. The start of the Arab Spring marks a change in the reporting style of contentious politics. The MENA media finds new topics to use (e.g. *Libya*), lets go of old preferences (e.g. *Iraq*) and only few topics maintain their popularity throughout the entire period of investigation (e.g. *violent protest*). We find that the use of most topics is not entirely a decision made by the newspaper, the state also having a say. Topics such as *revolution* or *Syria* seem to be objects of censorship, their prevalence being almost nil in countries with the lowest levels of press freedom. At the same time press freedom has no influence with regards to topics such as *political implications* or *democracy*. One explanation for this discrepancy might be that the latter topics do not exclude the current ruling regime from the process of change and the demanded action can be localized as sometime in the future as opposed to right now. As noted above, topic prevalence also depends on the intensity of foreign protests and this is true for almost all topics. Intensifying domestic protests always have the same effect on topic prevalence as the foreign protests, though in some cases this effect is not as pronounced. This also holds true for all but one topic (i.e. *Libya*) when we look at the influence of increased domestic censorship practices. The fact that this variable has a positive effect on topics such as *revolution*, which is one of the most evident objects of censorship among the autocratic MENA states, seems to imply that Arab journalists do fight back.

Our analysis has also a few limitations, the main consisting of the fact that we look only at the English-language publications, which might address a different segment of the population than the Arab-language ones. However, we consider that such reporting bias should be even more pronounced in the Arab-language publications. Other aspects



we consider to improve on are restricting censorship events to solely contentious politics and domestic media and improving on the topic model with as sentiment analysis.

## References

- Alkazemi, M. F., & Wanta, W. (2015). "Kuwaiti political cartoons during the Arab Spring: Agenda setting and self-censorship." *Journalism* 16(5): 630-653.
- Bairrett, Richard L. (2015). "Executive Power and Media Freedom in Central and Eastern Europe." *Comparative Political Studies* 48: 1260-1292.
- Baum, M. A., & Zhukov, Y. M. (2015). "Filtering revolution Reporting bias in international newspaper coverage of the Libyan civil war." *Journal of Peace Research* 52(3): 384-400.
- Besley, T. J., and A. Prat (2006). "Handcuffs for the grabbing hand? Media capture and government accountability." *American Economic Review* 96: 720-736.
- Bhattacharyya, S., and R. Hodler (2015). "Media freedom and democracy in the fight against corruption." *European Journal of Political Economy* 39: 13-24.
- Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). "Latent Dirichlet Allocation". *Journal of Machine Learning Research* 3: 993-1022.
- Crabtree, C., D. Darmofal and H. L. Kern (2015). A spatial analysis of the impact of West German television on protest mobilization during the East German revolution. *Journal of Peace Research* 52: 269-284.
- Collingwood, L. and J. Wilkerson (2012). „Tradeoffs in accuracy and efficiency in supervised learning methods“. *Journal of Information Technology & Politics*, 9(3):298–318.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In Kittler, J. and F. Roli (Eds.), *Multiple classifier systems*. New York: Springer. 1-15.
- Dolezal, M., S. Hutter and B. Wueest (2012). Exploring the new cleavage across arenas and public debates: design and methods. In Kriesi, H., E. Grande, M. Dolezal, M. Helbling, D. Hoegliger, S. Hutter and B. Wueest, *Political Conflict in Western Europe*. Cambridge, MA: Cambridge University Press. 36-66.
- Entman, R. M. (2004). *Projections of Power: Framing News, Public Opinion, and U.S. Foreign Policy*. Chicago, IL: University of Chicago Press.
- Gehlbach, S. and K. Sonin (2014). "Government control of the media." *Journal of Public Economics* 118: 163-171.
- Gohdes, A. R. (2015). "Pulling the plug Network disruptions and violence in civil conflict." *Journal of Peace Research* 52(3): 352-367.
- Granovetter, M. (1978). "Threshold models of collective behavior". *American Journal of Sociology* 83: 1420–1443.
- Greene and Cross (2015). "Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach." Paper presented at the 5th annual conference of the European Political Science Association in Vienna, June 25-27, 2015.
- Hamdy, N. and Gomaa, E. H. (2012). "Framing the Egyptian uprising in Arabic language newspapers and social media." *Journal of Communication* 62(2): 195-211.
- Kim, H. J. Whitten-Woodring, and P. James (2014). "The Role of Media in the Repression-Protest Nexus: A Game-theoretic Model" *Journal of Conflict Resolution* 59: 1017-1042.

- King, G., J. Pan, and M. E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression," *American Political Science Review* 107: 1-18.
- Kuran, T. (1991). "Now out of never: The element of surprise in the East European revolution of 1989." *World Politics* 44: 7–48.
- Lohmann, S. (1993). "A signaling model of informative and manipulative political action." *American Political Science Review* 87: 319–333.
- Long, J. S. and Freese, J. 2006. *Regression Models for Categorical Dependent Variables Using Stata*, College Station, TX: Stata Press.
- Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013). "Efficient Estimation of Word Representations in Vector Space," CoRR, abs/1301.3781.
- Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. Gadarian, B. Albertson and D. Rand (2014). "Structural topic models for open-ended survey responses." *American Journal of Political Science* 58: 1064-1082.
- Weidmann, N.B. and E.G. Rød. 2014. "Making Uncertainty Explicit: Separating Reports and Events in the Coding of Violence and Contention." *Journal of Peace Research* Forthcoming.
- World Association of Newspapers (WAN-IFRA). *World Press Trends 2010* (2010) Darmstadt and Paris.

## Appendix

Table A.1: *Media data overview*

<b>Newspaper</b>	<b>Ownership</b>	<b>Country</b>	<b>N Paragraphs Total</b>	<b>N Paragraphs Relevant</b>	<b>Start of Coverage</b>	<b>End of Coverage</b>
BNA	State	Bahrain	1824	283	01.09.2009	31.12.2011
Gulf Daily News	Other	Bahrain	29788	2037	01.01.2009	31.12.2011
Daily News Egypt	Other	Egypt	56917	11880	01.01.2009	31.12.2011
The Egyptian Gazette	State	Egypt	3405	855	01.07.2010	31.12.2011
NINA	State	Iraq	335	42	01.01.2009	31.12.2011
The Kurdish Globe	Other	Iraq	2223	200	01.01.2009	31.12.2011
AmmanNet	Other	Jordan	579	61	01.07.2010	31.12.2011
The Jordan Times	State	Jordan	20347	3408	01.01.2009	31.12.2011
The Star	Other	Jordan	5344	413	01.03.2009	31.09.2011
The Daily Star	Other	Lebanon	51202	13219	01.01.2009	31.12.2011
JANA	State	Libya	1297	311	01.10.2009	31.12.2011
The Tripoli Post	State	Libya	3635	1574	01.01.2009	31.12.2011
Oman Tribune	Other	Oman	65985	16162	01.01.2009	31.12.2011
ONA	State	Oman	596	16	01.09.2009	31.12.2011
The Week	Other	Oman	706	3	01.01.2009	31.12.2011
Times of Oman	Other	Oman	31305	5898	01.01.2009	31.12.2011
The Palestine Chronicle	Other	Palestine	18722	1414	01.01.2009	31.12.2011
WAFSA	State	Palestine	3833	854	01.09.2009	31.12.2011
Asharq Alawsat	State	Pan-Arab	3958	1233	01.07.2010	31.12.2011
Dar Al Hayat	State	Pan-Arab	9383	1508	01.01.2009	31.12.2011
The Majalla	Other	Pan-Arab	1780	495	01.04.2009	31.12.2011
The Middle East Reporter	Other	Pan-Arab	13982	8025	01.08.2010	31.12.2011
Qatar Tribune	Other	Qatar	2593	251	01.09.2009	31.12.2011
The Peninsula	Other	Qatar	8928	4717	01.04.2010	31.12.2011
Arab News	Other	Saudi Arabia	8901	1517	01.01.2009	31.12.2011
Khaleej Times	Other	Saudi Arabia	14242	618	01.01.2009	31.12.2011
Saudi Gazette	Other	Saudi Arabia	36671	6726	01.01.2009	31.12.2011
7 Days	Other	UAE	6025	430	01.01.2009	31.12.2011
Gulf News	Other	UAE	34732	4104	01.01.2009	31.12.2011
The National	State	UAE	97884	11216	01.01.2009	31.12.2011
UMCI News	Other	UAE	3292	1274	01.01.2009	31.12.2011
SABA	State	Yemen	1660	775	01.07.2009	31.12.2011
Yemen Times	Other	Yemen	7506	2664	01.01.2009	31.12.2011

Table A.2: Results of negative binominal regressions on number of paragraphs covering contentious mobilization

	Model 1.0	Model 1.1	Model 2.0	Model 2.1	Model 3.0	Model 3.1	Model 4.0	Model 4.1
<b>Economic Control</b>	-0.637 (0.071) ***	-0.688 (0.070) ***	-0.631 (0.071) ***	-0.677 (0.070) ***	-0.636 (0.071) ***	-0.695 (0.070) ***	-0.633 (0.071) ***	-0.685 (0.070) ****
<b>Political Control</b>	0.111 (0.038) **	0.077 (0.037) *	0.110 (0.038) **	0.076 (0.037) *	0.200 (0.038) ***	0.078 (0.037) *	0.112 (0.038) **	0.070 (0.038)
<b>Legal Control</b>	0.335 (0.076) ***	0.368 (0.075) ***	0.334 (0.076) ***	0.366 (0.075) ***	0.344 (0.076) ***	0.369 (0.075) ***	0.330 (0.076) ***	0.371 (0.075) ***
<b>Owner (State)</b>	-0.426 (0.068) ***	-0.080 (0.078)	-0.421 (0.068) ***	-0.087 (0.079)	0.130 (0.068)	-0.074 (0.079)	-0.430 (0.068) ***	-0.123 (0.078)
<b>MENA Protests</b>	-0.111 (0.044) *	-0.072 (0.054)	-0.105 (0.044) *	-0.067 (0.054)	-0.129 (0.044) **	-0.071 (0.054)	-0.111 (0.044) *	-0.072 (0.054)
<b>Domestic Protests</b>	0.402 (0.097) ***	0.367 (0.120) **	0.364 (0.097) ***	0.366 (0.121) **	0.373 (0.097) ***	0.381 (0.121) **	0.403 (0.097) ***	0.373 (0.120) **
<b>Owner (State) MENA Protests</b>		-0.290 (0.084) ***		-0.302 (0.084) ***		-0.286 (0.084) ***		-0.302 (0.084) ***
<b>Owner (State) x Domestic Prot.</b>		0.410 (0.171) *		0.323 (0.171)		0.270 (0.173)		0.408 (0.172) *
<b>Censorships</b>	-0.039 (0.046)	0.161 (0.060) **						
<b>Owner (State) x Censorships</b>		-0.438 (0.074) ***						
<b>Suppression</b>			0.064 (0.050)	0.160 (0.073) *				
<b>Retaliation</b>			-0.136 (0.064) *	0.009 (0.081)				
<b>Owner (State) x Suppression</b>				-0.208 (0.097) *				
<b>Owner (State) x Retaliation</b>				-0.279 (0.113) *				
<b>Violent</b>					0.070 (0.056)	0.102 (0.075)		
<b>Non Violent</b>					-0.149 (0.059) *	0.076 (0.079)		
<b>Owner (State) x Violent</b>						-0.032 (0.103)		
<b>Owner (State) x Non Violent</b>						-0.479 (0.109) ***		
<b>Individual</b>							-0.046 (0.049)	0.126 (0.068)
<b>Organization</b>							0.038 (0.079)	0.139 (0.113)
<b>Owner (State) x Individual</b>								-0.351 (0.087) ***
<b>Owner (State) x Organization</b>								-0.195 (0.155)
<b>Log-Likelihood Ratio Test for Overdispersion (Model 1.0)</b>								133712 ***(df=22)
<b>AIC</b>	30629	30525	30627	30543	30634	30539	30632	30544

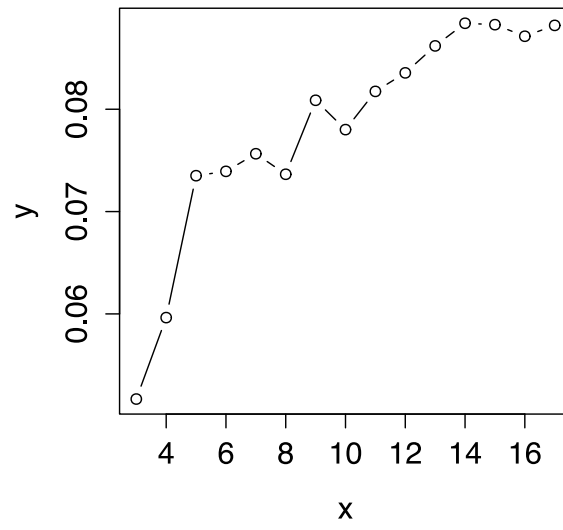


Figure A.1: *Results of word2vec topic coherence evaluation*

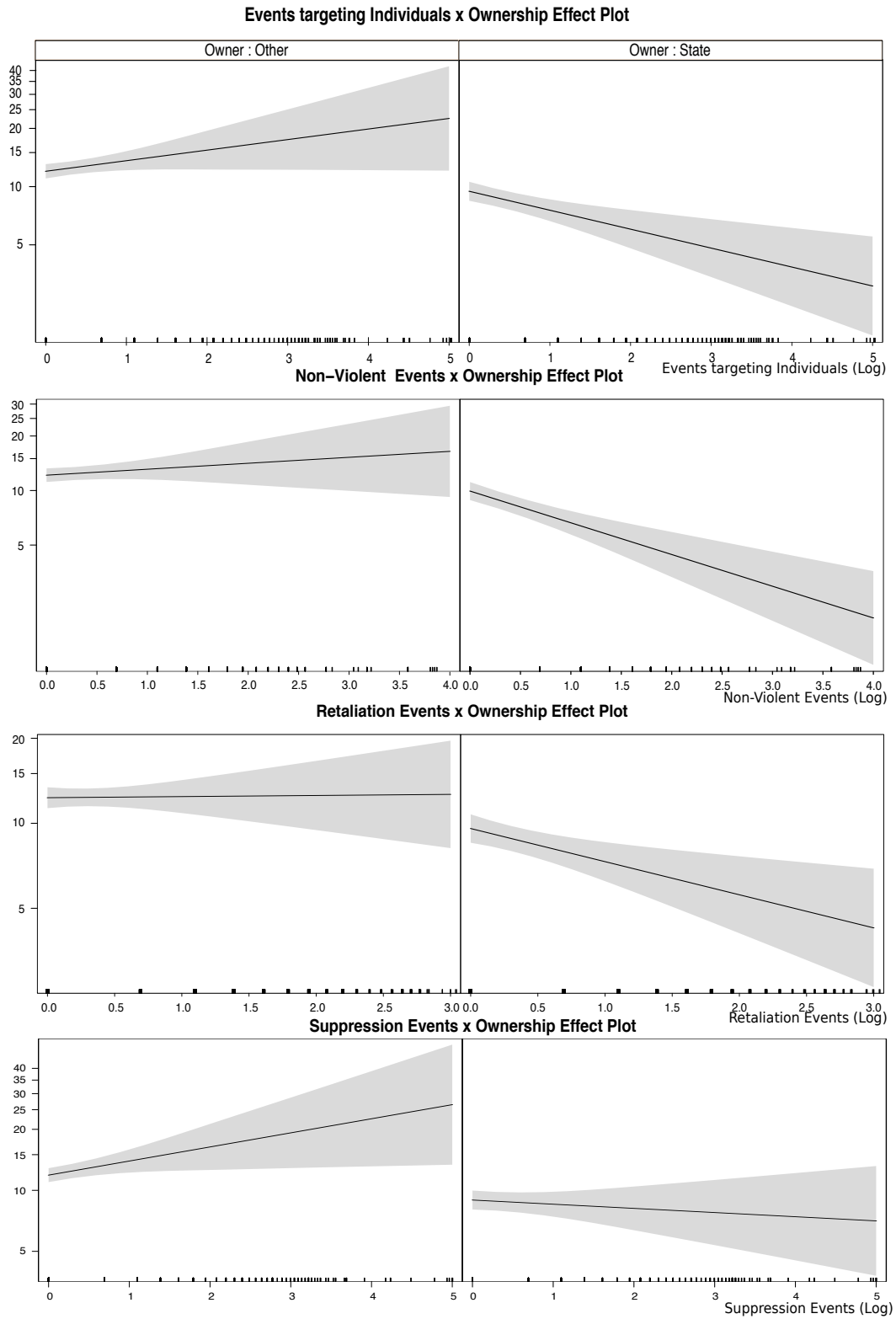


Figure A.2: Reporting bias of state-owned newspapers based on type of censorship employed