Moment of Madness: Shifting Sentiment and the Dynamics of Revolution in the Finnish Civil War

Teppo Lindfors*

February 24, 2025

Abstract

How do revolutions take hold? This paper investigates how a sudden shift in collective sentiment, triggered by the 1917 Russian Revolution, drove Finland into civil war. This change was made possible by two key factors: the sudden end of imperial censorship and the diffusion of revolutionary ideas from Petrograd. Analyzing millions of newspaper articles, I show that the removal of media restrictions, combined with the echoes of the Bolshevik revolution, led to a sharp rise in public opposition. Employing difference-in-differences and spatial regression discontinuity designs, I demonstrate that this transformation in public discourse played a critical role in mobilizing the Finnish population into an armed rebellion within ten months. Drawing on newly digitized interrogation records from 46,146 rebels, I find that participation was primarily driven by shifting beliefs about the likelihood of revolutionary success and reinforced by peer influence—over 40% of rebels cited social pressure, coercion, or communal expectations as key motivations. At the same time, government repression had a strong demobilizing effect, with public opposition unable to offset its impact, suggesting that while belief shifts can spark rebellion, they are insufficient to sustain it. These results underscore the role of information shocks and social contagion in triggering collective action, revealing how quickly political stability can unravel when suppressed grievances meet a unifying narrative.

Keywords: censorship, collective action, political violence

JEL Codes: D74, D83, N44

*Department of Economics, Lund University, teppo.lindfors@nek.lu.se. I am indebted to Gunes Gokmen and Matti Mitrunen for their valuable support and advise. I thank Konrad Burchardi, Walker Hanlon, Leander Heldring, Torben Johansen, Andrei Markevich, Joel Mokyr, Erik Mohlin, Elisa Mougin, Pauline Rossi, Valeria Rueda, Sakari Saaritsa, Mohamed Saleh, Petros Sekeris, Alexander Yarkin and all the seminar participants at University of Helsinki, Northwestern, Lund, Aarhus and Arne Ryde Workshop on Culture, Institutions and Development for their helpful comments and suggestions. I gratefully acknowledge financial support from Yrjö Jahnsson Foundation.

1 Introduction

Revolutions are paradoxical events: they are seemingly unpredictable yet appear almost inevitable in retrospect (Kuran 1989, 1991). Societies that seem stable and docile can suddenly erupt into mass protests, overturning regimes in a matter of weeks. There is a critical, underexplored element in this process – the rapid transformation of public sentiment. Revolutions are "moments of madness", when societies swiftly cross the threshold from discontent to action, creating a new reality that was previously unthinkable (Zolberg 1972).

A central mechanism in this transformation is the autocratic regime's control over media and public discourse. By censoring dissent and controlling the narrative, authoritarian governments maintain an image of stability. However, when this control breaks down, it opens the floodgates of opposition sentiment, altering the tone and content of public discourse. Can changes in media control and collective sentiment push societies over the edge?

This paper exploits the dissolution of autocratic Russian Empire to study the effect of rising public opposition on mass mobilization. The context is Finland, the Empire's hinterland. Before becoming independent in December 1917, Grand Duchy of Finland was an autonomous part of Imperial Russia, and subject to severe state censorship, imposed by the Russian authorities. When Russian Tsar Nicholas II was overthrown in February revolution, the censorship was abolished overnight, triggering a dramatic surge in newspaper opposition and a bandwagon effect as revolutionary ideas rapidly diffused. I utilize this historical episode as a natural experiment to quantitatively assess the political implications of media deregulation and a swift shift in public narrative.

The purpose of imperial censorship was to uptain the stability of Russian Empire by preventing the dissemination of texts that "incited hatred against the government or general unrest" (Kuusanmäki 1980). When censorship was removed, newspapers brought widespread dissent into spotlight. In the following months of 1917, the Finns could read stories on prevalent rioting in Petrograd or Helsinki, separatist essays, visions of socialist revolution or agitation against the bourgeoisie, themes which had been completely taboo for the past years.

Russian Revolution of 1917 is one of the most consequential historical episodes of the 20$^{\text{th}}$ century. The revolution inspired regime changes across all inhabitated continents. Social historian Eric Hobsbawm does not skimp superlatives when describing its historical significance: *"The October revolution produced by far the most formidable revolutionary movement in modern history. Its global expansion has no parallel since the conquests of Islam in its first century. A mere thirty to forty years after Lenin's arrival at the Finland Station in Petrograd, one third of* *humanity found itself living under regimes directly derived from the 'Ten Days That Shook the World' (Reed 1919), and Lenin's organizational model, the Communist Party."* (Hobsbawm 1995, p. 55) Another purpose of this paper is to shed light on the repercussions of Russian Revolution, and the spatial spread of revolutionary movements overall.

I find that an increase in pro-opposition rhetoric encouraged people to mobilize in Finnish Civil War, fought 10 months after the removal of censorship. The civil war was a power struggle in a newly independent state, also referred to as "fringe phenomenon" of the Russian civil war (Rasila 1969). According to qualitative and anecdotal evidence, newspaper agitation and the precedent set by the Russian Bolsheviks enticed the Finnish working class to join in paramilitary organizations, and in January 1918 rebel against the bourgeois government. My study is the first one to quantitatively assess this historical conjecture.

My empirical analysis builds on two distinct identification strategies. For one, in my baseline analysis I utilize local difference-in-differences setup with a shift-share exposure to a change in public opposition. In these regressions, the exposure is determined by baseline journal composition. For two, I exploit spatial fuzzy regression discontinuity design, taking advantage of the fact that target audience of local newspapers is often province-based. I restrict my focus on the area around Viipuri province border. As a border region to the then capital of Russia, Petrograd, Viipuri province was under exceptionally strict censorship but also heightened exposure to the spread of revolutionary ideas. The shift in public opposition after the termination of censorship was accordingly stark there in comparison to the neighboring provinces.

To measure the effect of growing public opposition, I construct a novel historical dataset, combining quantitative and qualitative data sources. To quantify public opposition, I collect the universe of some 9 million Finnish newspaper articles from January 1, 1916 to January 26, 1918. I conceptualize public opposition by the usage of inflammatory words, such as "revolution", "democracy" and "freedom". To map the shift in public opposition to regional level, I compose an extensive geographical breakdown of newspaper subscriptions in Finland, making use of previous research, archival sources and local history books. This breakdown covers around 70% of the total newspaper circulation at the time.

Figure 2 illustrates the evolution of public opposition in the newspapers. The graph demonstrates, that the removal of censorship corresponds to a dramatic shift in pro-opposition rhetoric. Overall, the usage of inflammatory words grows over five-fold.

What makes the Finnish Civil War a particularly compelling revolution to study is the detailed data on rebel motives. Over the three months immediately following the conflict, victorious bourgeois Whites interrogated all $67,703$ captive Reds, asking each insurgent about their personal motivation to join the rebel force. These interrogations were typed up and combined with other material on the accused, such as personal information and testimonies, to form treason court archives, a vast collection of 750,208 archival documents. I build a tailored deep learning pipeline to recognize and digitize interrogation records among this archival material. This data offers a rare vantage point to investigate revolutionaries' self-perceived motivations to rebel, and how they might be affected with growing public dissent.

I measure mobilization with various different variables. As my main outcome, I use war mortality rate, an indicator of extreme dissent. I complement the metric with alternative proxies, including overall mortality rate, the foundation of Red Guards, the volume of post-conflict investigations and the number of strikers, finding similar results. For the spatial regression discontinuity design, I build a unique village-level dataset for villages in proximity of the Viipuri province border. To do this, I trace the coordinates of 1,500 historical villages, drawing from a toponomastic database NameSampo and a number of online map interfaces, such as Maps of Karelia, GeoHack and Google Maps.

I show that a one-word increase in exposure to inflammatory rhetoric per every 100 articles is associated with a 20-log point increase in war mortality. This estimate would imply that increasing inflammatory rhetoric by one standard deviation, 0.88 words, would have caused war mortality to increase by 18 log points. To mitigate endogeneity concerns, I demonstrate that the more exposed municipalities were comparable in baseline characteristics and preexposure trends. Complementarily, I establish that regression discontinuity design produces qualitatively and numerically very similar results. Both designs are robust to a battery of sensitivity tests, including matching on observables, omitting the province of Viipuri, different RD polynomials or various bandwidth lengths.

My estimates indicate that the shift in public opposition only affected the mobilization of the insurgents, not the Senate-led Whites. The regions which experienced a greater increase in public opposition had a higher fraction of Red casualties and a higher fraction of Red Guard casualties in the civil war. These regions also had a higher strike activity in 1917 and more post-war investigations once the hostilities ended. Using survival analysis, I further demonstrate that an exposure to a greater increase in pro-opposition rhetoric resulted in a greater risk to founding a Red Guard, a paramilitary organization facilitating the collective action of the insurgents.

My findings suggest that the increase in public opposition can be attributed to at least two contingent sources. I find that pro-opposition rhetoric grew more in newspapers and regions which were exposed to more stringent censorship, or more extensive coverage of the Russian Revolution. My interpretation of these results is, that both disruptions to the enforcement of censorship or revolutionary examples elsewhere can contribute to revolutionary movements. These findings shed light on the contagious nature of popular uprisings and the critical role of media in their spread.

I argue that public opposition drives mobilization primarily by shifting people's beliefs about the likelihood of a successful revolution. This interpretation is supported by two key findings. On the one hand, the stated motivations of rebels remain similar across municipalities, regardless of the level of public dissent, yet insurgent participation is consistently higher in areas with stronger opposition. This suggests that dissent does not alter individuals' intrinsic reasons for rebellion but instead acts as a rallying cry for action. On the other hand, strategic complementarity plays a crucial role—individuals are more likely to mobilize when their peers do, and this effect is particularly strong where public opposition is more explicit.

Beyond shifting beliefs, peer influence itself emerges as a key driver of mobilization. Newly digitized interrogation records from 46,146 rebels reveal that the most commonly cited motivation for participation was social pressure—41% of rebels pointed to encouragement, coercion, or communal expectations as primary factors in their decision to take up arms. These results highlight how sudden shifts in public information, reinforced by social contagion, can rapidly turn passive dissent into coordinated rebellion.

To better understand the interrelationship between dissent and mobilization in different contexts, I provide several additional results. First, I discover that public opposition had a particularly strong mobilizing effect in municipalities with a strong socialist voter base, high levels of economic inequality, and the presence of workers' houses—factors that likely facilitated coordination and reinforced collective grievances. Second, the onset of government repression led to a sharp decline in mobilization, which public dissent was unable to offset, suggesting that while shifts in beliefs can trigger rebellion, they are insufficient to sustain it in the face of targeted crackdowns.

My article joins a growing literature on the effects of traditional media on various political variables (Enikolopov, Petrova, and Zhuravskaya 2011; Yanagizawa-Drott 2014; Adena et al. 2015; Bai, Jia, and Wang 2023). I complement this work in three ways. First, to my knowledge, I am the first to focus on a historical episode where the enforcement of state censorship abruptly deteriorates. Considering the potential of new technology to circumvent the censorship of traditional media (Manacorda and Tesei 2020; Enikolopov, Makarin, and Petrova 2020; Guriev, Melnikov, and Zhuravskaya 2021), this is relevant on several angles. Second, the scale of my setting is the universe of newspaper media, making it appropriate to investigate the change in the society's popular narrative. Third, I explore a famous historical instance where revolutionary ideas spread from one country to another through the news, unveiling details of revolutionary bandwagons (Kuran 1991; Aidt and Jensen 2014; Acemoglu, Naidu, et al. 2019).

This paper also substantiates the large body of theoretical scholarship on collective action in revolutions (Kuran 1989; Lohmann 1994; Vives 2005; Angeletos, Hellwig, and Pavan 2007; De Mesquita 2010; Edmond 2013). The results are consistent with the general idea of collective action games characterized by strategic complementarity, that changes in prior beliefs on the strength of the status quo may inflict insurgency.

Lastly, this paper further contributes the research on Finnish Civil War. Until the 1960s, newspaper agitation was among the most cited explanations for fuelling the insurgency (Aminoff 1918; Hannula and Korhonen 1956)[1]. Since then, the research stressed structural factors – such as class antagonism – as an important precondition behind the conflict (Alapuro 2018; Haapala 1995; Meriläinen, Mitrunen, and Virkola 2022). In recent years, the agitation-hypothesis has regained popularity (Ihalainen 2017; Matikainen 2018; Turunen 2021). I contribute to the literature by assessing the agitation-hypothesis using a plausibly causal research design.

The rest of the paper is structured as follows. Section 2 outlines the historical background of Finnish Civil War, and the preceding collapse of imperial censorship. Section 3 describes the data. Section 4 discloses the empirical strategy for testing the relationship between public opposition and mobilization, while section 5 presents the results. Section 6 explores the different mechanisms behind the surging opposition. Section 7 assesses robustness of the results. Section 8 concludes.

2 Historical Background

> *"The working class does not believe in God, but they believe in the Devil, whom they call the bourgeois."*
>
> – In *Uusi Päivä* -newspaper, May 3, 1918

The economic conditions in the Grand Duchy of Finland were ripe for a rebellion during the decade preceding the civil war. The society was extremely unequal. Gini coefficient for income inequality was as high as 62 in 1910 (Meriläinen, Mitrunen,

[1] For a comprehensive survey on the historical research of Finnish Civil War, see Tikka (2018).

and Virkola 2022). Tenant farming was common: almost 50% of the farmers were renting their land. Indeed, the ownership over land was the most salient motive for class antagonism in the countryside. As Finland was still a relatively agrarian economy, industrial jobs were scarce, which in combination with a fast population growth resulted in rural overpopulation. The reduction in land to labour ratio, at least for high quality soil, manifested itself in downward mobility, particularly among those who would during Winter 1918 join the insurgency (Santavirta and Stuhler 2020).

To make matters worse, the economy was hit by a number of shocks in 1917. GDP contracted by astonishing 16%, a single largest reduction in the economic history of Finland. Unemployment soared. Russian Revolution disrupted the importation of grain, which the food supply system was dependent on. Attempts to ration food consumption failed, while black markets flourished, and shortages emerged. Prices doubled, partly due to the global inflation wave following World War I (Eichengreen 1992). The deteriorating living conditions lead to growing dissent, triggering riots and protests. If deprivation was not enough to spark a revolution in itself, newspapers were not shy to point out scapegoats: in papers with left-leaning editors, the devils were the profiteering bourgeoisie and the grain hoarding landowners, while for right-leaning editors, the blame fell on rioting socialists.

In March 1917, the Tsarist Regime in Russia ended, as Tsar Nicholas II was deposed in February revolution. The turmoil created a power vacuum in the Grand Duchy of Finland. Government was under confusion, and the abolition of the local law enforcement, Russian Gendarmes, incentivised left- and right-wing groups to found their own paramilitary organizations, Red and White Guards (Suodenjoki and Turunen 2017). Yet, Finland was still a part of Russia. Strikes, which were deemed illegal by the Empire during the World War, became a habitual way of protesting: some 150,000 workers took part in striking during 1917. Kirby (1978) describes the situation aptly: *"Finland was thus plunged into a situation in which full authority was exercised by none and contested by all."*

Political deadlock aggravated the frustrations of the working class. Despite being the most popular labor-oriented party in Europe, the Finnish social democratic party (SDP) was unable to achieve its goals while Finland was part of the Empire, because the Tsar routinely vetoed reforms (Upton 1980). In spring 1917, the party's most important objectives were an eight-hour working day, a land reform and establishing municipal democracy (Alapuro 2018).

The dethroning of Tsar provided the social democrats a new momentum for reforms. In July 1917, SDP-led parliament approved the so called Power Act, a declaration of home rule in Finland, as a first leap toward independence. Attempting to prevent the Empire from collapse, the Russian Provisional Government effectively barred the Power Act in a Tsarist fashion by dissolving the Finnish parliament. In October, SDP lost the following elections, and was now overpowered by the bourgeoisie both in parliament and in the executive organ, Senate. The mood amongst social democrats and their supporters was, that bourgeois politicians had together with the Russians co-orchestrated the dissolution of parliament, and essentially ousted them from positions of power (Lindfors, Meriläinen, and Mitrunen 2024). The bourgeois party-leaders were viewed as betrayers, and their mandate to rule was considered illegitimate (Siltala 2023).

Towards the end of 1917, the social conditions for a revolution were in place. All one needed was a spark, which appeared in the form of Bolshevik Revolution of October 1917. A week after, the socialists launched a general strike in Finland, which together with street demonstrations resulted in the bourgeois parliament to approve the eight-hour workday and universal suffrage in local elections. These concessions could not hold the storm. In the following months, the militarization of Red and White Guards intensified, after culminating to a full conflict in January 1918.

Finnish Civil War was a short and bloody struggle over power in a newly independent state between the bourgeois Whites and the socialist Reds, following the collapse of the Russian Empire. It started out in late January, 1918, and lasted about four months until the definitive victory of Whites in May, 1918. The total death toll of the war was around 36,000 casualties or 1.6% of the population, of which 27,000 Reds.

2.1 Public Opposition in the Newspapers

According to the contemporaries, especially among the victors of the civil war, inflammatory newspaper rhetoric was a key component in enticing the working class to take up arms (Ryömä 1918). Indeed, a number of future leaders of the Reds wrote polemic writings in newspapers during 1917, particularly in leftist journals. Based on the rebels' narrative, the root cause of all social wrongs, whether unemployment, hunger or sexual harrassment, was the system and by extension, the current establishment (Turunen 2021; Upton 1980). The most colourful arguments juxtaposed agitation with mass suggestion, driving gullible countrymen to unimaginable atrocities (Åström 1918). In his diary, novelist Juhani Aho wrote

> *"It is hypnosis and intoxication, which I deduce from the fact that it has also taken hold of those who have not been in direct contact with misery, but are carried along, inspired solely by the ideology and words."*
>
> – In *Hajamietteitä kapinaviikoilta*, January 28, 1918

Recent historical work supports the argument that the newspapers had a pivotal role in fuelling the rebellion (Ihalainen 2017; Suodenjoki and Turunen 2017; Matikainen 2018).

Until the Spring of 1917, public opposition was firmly bottled by wartime censorship, imposed by the Governor-General's decree in 1914. The decree prohibited "the dissemination of writings and communications that incite hatred against the government or general unrest, as well as the glorification of crimes and the distribution of confiscated printed materials". In practice, any news that presented the Empire in an unfavorable light could be plausibly censored. Typical examples of forbidden topics included Russification policies, street demonstrations or stories shedding light on social grievances (Leino-Kaukiainen 1984). While foreign newspapers were not censored, their importation was restricted. For instance, numerous Swedish journals and all German or Austrian papers got blacklisted (Kuusanmäki 1980).

Censorship was not the only instrument Imperial Russia exploited to curtain opposition. Over World War I, freedom of assembly was restricted and strikes were completely illegalized. The objective of Russian authorities was first to prevent and second to marginalize revolutionary action. In addition to the censors, political mindscape was monitored by the Russian Gendarmes, secret police and a network of espionages (Tommila, Nygård, and Salokangas 1987).

Vocal opponents of the system got punished. Journals got their issues confiscated, paid large fines and were under a constant threat of closure. The high costs of public opposition resulted in preventive censorship by the editors (Kuusanmäki 1980). What is more, government officials got fired and the most prominent dissidents were either exiled or imprisoned. In a famous instance, a member of the parliament Pehr Svinhufvud was banished to Siberia in 1914 for refusing to acknowledge Konstantin Kazanski, a Russian citizen, as the Finnish attorney general (Upton 1980).

All news articles that the censors surveyed and approved were labelled. Figure 2 shows the evolution censorship labels per page from 1900 to 1930. It vividly demonstrates the disruption censorship caused in the newspapers in 1914–1917. At the height of wartime censorship in 1916, over half the pages had at least one surveyed article. Censorship was total, and it had a strong effect on public sentiment that the newspapers upkept. The policy silenced socialist rhetoric and agitation altogether. Turunen (2021) shows, that words such as "freedom" and "Russki" vanished from the newspaper corpus completely.

Following the Empire's collapse, censorship was abolished 20$^{\text{th}}$ of March, 1917. Figure A1 displays the subsequent upturn in public opposition in detail. Overnight, inflammatory words such as "revolution", "democracy" or "freedom" swarmed the newspapers. The frailty of Russian rule and confusion regarding local authority became evident all across the country.

Another surge in public opposition occured in November 1917, inspired by the Bolshevik Revolution. According to historical research, Russian bolsheviks explicitly encouraged Finnish socialists to revolt (Kirby 1978; Salkola 1985). Many took this request to heart, as political violence escalated. Akin to the bolsheviks, the aspiration of newly founded Red Guards was to "tear down the old world, and create a new, better one in its place". Around the same time, newspaper rhetoric took a darker tone. For instance, usage of the word *lahtari* (butcher) – a derogatory nickname for the bourgeoisie – exploded in the press (Figure A1). Turunen (2021) describes the working class' ambiance towards the end of 1917 as follows: *"'No beast is as cruel as the bourgeois whose purse is touched.' In the final stage of this evolution, during November and December, the beast began to pursue not only money for its bottomless purse but also proletarian blood, as the economically-motivated bourgeoisie became the militaristic butcher in the socialist imagination."*

3 Data

I combine a variety of quantitative and qualitative data to examine the interrelation between public opposition and mobilization. My primary measure for mobilization, war mortality, I construct using open data from WarVictimSampo 1914–1922. WarVictimSampo documents all war-related deaths in Finland from 1914 to 1922. To quantify public opposition, I exploit the universe of newspaper articles in Finland from two years prior the outset of the civil war. The resulting unbalanced panel dataset covers mortality, the change in public opposition and related covariates from 1914 to 1922 from around 500 municipalities.

Public Opposition in the Newspapers To measure the change in public opposition in Finland around the abolition of censorship, I use the universe of newspaper articles from 1st of January, 1916 to 26th of January, 1918. This data was OCRed and made publicy available by National Library of Finland. I scrape the articles. In total, this data covers 9.3 million newspaper articles from 163 journals.

Following previous studies applying newspaper data (Baker, Bloom, and Davis 2016; Bai, Jia, and Wang 2023; Venturini 2023), I conceptualize public opposition as the usage of inflammatory words from Figure A1, including "revolution", "democracy", "freedom", "oppression", "anarchy" and "butcher". I choose this set of words building on Turunen (2021), who uses them to document and describe the change in newspaper rhetoric in Finland after the removal of censorship. I then compute the average term frequency, ATF, for this group of words by journal, and define the change in public opposition at the journal-level as

$$\Delta PublicOpposition_d = ATF_{d,1} - ATF_{d,0} = \frac{z_{d,1}}{M_{d,1}} \cdot 100 - \frac{z_{d,0}}{M_{d,0}} \cdot 100$$

where the subscripts d, 0 and 1 index journals and periods before and after the termination of censorship. $z_{d,t}$ is the raw count of inflammatory words in journal d in period t, and $M_{d,t}$ is the number of news articles in journal d in period t. I scale average term frequency by multiplying by 100, so that it presents words per 100 articles.

Censorship I measure censorship by exploiting a unique episode in Finnish history during World War I, when censorship was widespread and importantly, explicit: each news article read and approved by a censor was labelled with the phrase "S. H." (or "K. C."), which translates "approved by the censor".[2].

I operationalize the severity of censorship by journal by computing the fraction of articles carrying a censorship label.

Newspaper Subscribership Data To map the journal-level variables to municipal level, I collect newspaper subscribership data from Tommila, Raitio, and Aalto (1977). This data contains the geographical subscriber base of each of the major newspapers in Finland around early 20$^{\text{th}}$ century. I augment this data with subscribership records from archival sources and local history books to compile a relatively complete geographical breakdown of journal subscriptions in Finland. Overall, the volume of circulation in this breakdown covers around 70% of the total circulation in Finland at the time.

Outcomes The primary source for my outcome variables is WarVictimSampo 1914–1922, a dataset documenting the war victims in Finland in 1914–1922. The data consists of around 40,000 death records, majority of which from the Finnish Civil War. Before that, the records consist of voluntary fighters in WWI and the victims of violence over the prelude to the civil war. From 1919 onwards, most deaths are related to Finnish volunteers who died fighting in the Russian civil war. This data contains basic individual variables, such as name, age, marital

[2] In Finnish, *sensuurin hyväksymä*. In Swedish, *krigs censuren* (K. C.).

status, living municipality, living village and the municipality which the person deceased in, but also information on which side the person in question fought on and whether he had joined to a guard, among other things.

Interrogation Records To measure mechanisms, I exploit the complete interrogation records of captured Reds, which offer a rare access on detailed information describing the motivations of revolutionaries. To this end, I recognize and digitize the interrogation records among 750,000 treason court documents by developing a tailored multistep HTR pipeline. Further details on the digitization process are available on Appendix B.

The interrogation records consist of interrogation material produced as evidence to guide the convictions of imprisoned Reds in treason courts. Overall, 145 treason court departments were founded across Finland, each with five members: two scholars of law, one White officer and two "trustworthy citizens". The courts started their work only two weeks after the war had ended on May 15th of 1918, and continued to operate actively through the summer of 1918. With 75,575 captives, this amounts to sentencing 5 to 6 people each day (Kekkonen 1991).

Almost 90% of the accused got convicted with custodial sentence. For most, this implied 1 to 3 years of penitentiary, including forced labour for the state (*kuritushuonerangaistus*). The average sentence was four years in penitentiary. 555 captives were sentenced to death, but executions took place in prison camps also without court orders. Typically, the accused were sentenced on the basis of aid to treason.

The interrogation records contain details about the accused's occupation, name, birthplace, domicile, affiliation with a specific Red Guard department, the timing of their enlistment, and, most importantly, their personal account of the motives for joining the Red Guard. The material thus presents an unparalleled opportunity to investigate the underlying causes of social revolts, particularly how individuals' motivations to rebel are shaped by the perceived sentiments of others. An excerpt from the records is shown in Figure B2.

4 Empirical Strategy

My hypothesis is that the municipalities where the prior belief of opposition support increased the most, saw the largest increase in mobilization. I test this hypothesis by comparing war mortality between regions with differential change in public opposition in the newspapers. I quantify the change in public opposition in municipality m by computing a shift-share measure as the dot product of the baseline journal subscriber shares in the municipality and the journal-level change in inflammatory words

$$\Delta PublicOpposition_m = \sum_d w_{md} \Delta InflammatoryWords_d \tag{1}$$

where $w_{md} = S_{md}/S_m$ is the baseline subscriber share of journal d in municipality m. By focusing on the *change* in word frequency I am effectively able to measure the turn in pro-opposition rhetoric following the abolition of censorship, while purging out the baseline level of false positives.[3]

4.1 Difference-in-Differences at the Municipal Level

My baseline specification is the following difference-in-differences equation

$$y_{mt} = \alpha_m + \lambda_t + \beta \Delta PublicOpposition_m \times Post1917_t + \boldsymbol{x}'_m \boldsymbol{\gamma}_t + \varepsilon_{mt}, \tag{2}$$

where y_{mt} the outcome of interest in municipality m in year t, and $\Delta PublicOpposition_m$ is the average change in inflammatory words in the newspapers subscribed in municipality m, as described in equation (1). By including municipality fixed effects, α_m, I control for time-invariant unobserved differences across municipalities. Year fixed effects, λ_t, eliminate the impact of common changes in the outcome. $\boldsymbol{x}_m$ is a vector of municipality-level covariates, including geographic and demographic controls, as well as variables measuring prior mobilization and political orientation, all interacted with the $Post1917_t$-dummy. Standard errors are clustered at municipal level.

The coefficient of interest is β, which captures the causal effect of greater exposure to a change in public opposition. The geographical distribution of $\Delta PublicOpposition_m$ is shown in Figure 3.

The identifying assumption allowing causal interpretation is, that the average outcomes for the more and less exposed municipalities would have evolved in parallel absent the change in public opposition (Goldsmith-Pinkham, Sorkin, and Swift 2020). To test the validity of this assumption, Table 1 presents results of balance tests of a number of important covariates before the exposure. Imbalance would not necessarily invalidate my empirical strategy, unless correlated covariates can predict changes in the outcome if interest. Reassuringly, I find that the munici-

[3] In this context, false positives are news which are not really measuring public opposition, but for one reason or the other include at least one of the inflammatory words.

palities which experienced a greater change in public opposition did not differ in their preexposure characteristics, apart from two exceptions. The more exposed municipalities were less inclined to vote for social democrats and had less rugged terrain. Since both imbalances would predict a *lower* mobilization ex-ante, I conclude that the covariate balance provides support for the identifying assumption of parallel trends.

For a visual inspection of the parallel trends assumption, I estimate an event-study regression

$$y_{mt} = \alpha_m + \lambda_t + \sum_{\tau \neq -1} \beta_\tau \Delta PublicOpposition_m + \boldsymbol{x}'_m \boldsymbol{\gamma}_t + \varepsilon_{mt}, \tag{3}$$

which is identical to equation (2), apart from the fact that each variable is now interacted with year effects, not the $Post1917_t$-dummy.

To study heterogeneity and mechanisms, I use a triple-difference specification of the form

$$\begin{aligned} y_{mt} = \alpha_m + \lambda_t + \zeta \Delta PublicOpposition_m \times Post1917_t \times x_m + \\ \beta \Delta PublicOpposition_m \times Post1917_t + \boldsymbol{x}'_m \boldsymbol{\gamma}_t + \varepsilon_{mt}, \end{aligned} \tag{4}$$

where x_m is the variable of interest.

4.2 Fuzzy RDD at the Village-Level

A main concern that could potentially bias the difference-in-differences analysis is demand-based selection. If initially revolutionary regions were in higher demand of opposition-sided newspapers, the estimates would suffer from upward bias. This is arguably unlikely, as the outcomes measuring mobilization are on parallel trends prior the treatment, and I further control for political orientation and historical mobilization by municipality, which are expected to capture a lot of variation in demand. Nevertheless, I address the concern by exploiting a spatial regression discontinuity design, while utilizing a locality-based variation in newspaper demand, in contrast to taste-based demand.

My RDD identification strategy relies on the premise that local newspapers primarily target province-based audiences.[4] This implies that two households situated in neighboring villages across a provincial border are exposed to distinct news en-

[4] For a similar identification strategy, see Snyder Jr and Strömberg (2010).

vironments, shaped by different levels of public opposition. Figure A7 illustrates this pattern by showing how news consumption shifts abruptly at province borders, as local markets are dominated by province-specific content. Since provinces held minimal political authority at the time, neighboring villages across borders shared similar institutions, culture, and geography. This setting provides a credible framework for estimating the causal effect of news exposure using a border-based RDD.

The analysis focuses on Viipuri province, where shifts in newspaper rhetoric were more pronounced than in any other region of Finland, particularly relative to neighboring provinces such as Häme, Kuopio, Mikkeli, and Uusimaa. As a border region adjacent to Petrograd, Viipuri faced exceptionally strict censorship but also greater exposure to revolutionary ideas (Leino-Kaukiainen 1984).

The discontinuity of newspaper portfolio around province borders is not perfect, however, as local newspaper circulation does not align exactly with province borders, and national newspapers are read all across country. This is why I deploy a spatial fuzzy regression discontinuity design, where I instrument the local change in public opposition by an indicator variable, which gets value one for villages in Viipuri province, and is zero otherwise.

The estimation equations for the first and second stages are specified as the following system:

$$\Delta PublicOpposition_v = \pi + \kappa D_v + f(\text{location}_v) + \boldsymbol{x}_v'\boldsymbol{\delta} + \eta_v \tag{5}$$

$$y_v = \alpha + \beta \Delta \widehat{PublicOpposition}_v + f(\text{location}_v) + \boldsymbol{x}_v'\boldsymbol{\gamma} + \varepsilon_v \tag{6}$$

where $\Delta PublicOpposition_v$ represents the average change in inflammatory words in the newspapers subscribed in village v, D_v is an indicator variable equal to 1 if the village is located within Viipuri province and 0 otherwise, y_v is the outcome of interest in village v, $f(\text{location}_v)$ is the RD polynomial controlling for geographic location, $\boldsymbol{x}_v$ is a vector of village-level covariates, and η_v and ε_v are error terms. In the baseline specification, I use a local linear polynomial in longitude and latitude, with a sample of villages within 75 kilometres of the province border. I check robustness for several different forms of RD polynomial and bandwidths.

Figure 7 presents the first stage regression. It shows that the exposure to a change in public opposition experiences a clear discontinuity at the Viipuri province border. On average, the usage of inflammatory rhetoric increased by about 2 words (or 40%) more per every 100 articles in Viipuri province in comparison to the neighboring provinces.

The key identifying assumption of the regression discontinuity design is that all other factors affecting mobilization vary smoothly at the province border. To assess its plausibility, Figure A9 presents results of balance tests for the full set of controls from the baseline analysis. Consistent with the identifying assumption, I find balance with respect to all covariates.

5 Results

In this section, I present empirical evidence on the connection between the eruption of pro-opposition rhetoric from March 1917 and mortality in the Finnish Civil War of 1918. I start by studying the question using a local difference-in-differences setup. I proceed to address potential taste-based endogeneity of the treatment by applying a spatial regression discontinuity design. I report further analysis of the relationship by focusing on other mobilization outcomes than the aggregate conflict mortality, such as rebel mortality, the foundation of paramilitary organizations and striking activity. Last, I examine the heterogeneous effects of public opposition across different contexts.

5.1 Identification Using Difference-in-Differences

Figure 4 presents evolution of the event-study β_τ-coefficients from equation (3). The outcome in panel (a) is war mortality, while in panel (b) it is overall mortality. The estimates signify, that a greater increase in public opposition caused a sharp increase in mortality over the civil war of 1918. Specifically, increasing inflammatory rhetoric by one word per 100 articles increases war mortality by about 20 log points. The corresponding increase in overall mortality is 10 log points. Focusing on the more pertinent war mortality, these estimates would imply that increasing inflammatory rhetoric by one standard deviation, 0.88 words, would have caused war mortality to increase by 18 log points.

The pattern of estimates in Figure 4 demonstrate, that there were no differential pretends in outcomes in municipalities which were exposed to a greater increase in public opposition, relative to the municipalities which experienced a smaller increase. This observation provides support for the parallel trends assumption necessary for identifying a causal effect.

Table 2 reports baseline estimates from specification (2), separately for continuous and binary treatment. Binary treatment is an indicator variable, which gets value one for municipalities with above median change in public opposition, and is zero otherwise. These results reinforce the observations delineated in the event-study plots: a one-word increase inflammatory rhetoric per 100 articles induced a 22 log point increment in war mortality rate. In terms of overall mortality, the corresponding effect was about a half of that.

5.2 Identification Using Fuzzy RDD

Table 5 reports the second stage estimates from specification (6), while using log of war mortality as the dependent variable. Standard errors are clustered at the village level. The estimates in Panel A show a positive and significant effect of public opposition on war mortality. Overall, increasing inflammatory rhetoric by one word per 100 articles increases war mortality by 20 log points. These estimates are qualitatively and numerically very close to the ones found using difference-in-differences design, shown in Table 2.

Columns (1) to (6) demonstrate that the result is robust to various specifications. Column (1) specifies the RD polynomial as local linear polynomial in longitude and latitude. Column (2) uses a local linear polynomial in distance to the threshold, while column (3) uses both the coordinates and the distance. Column (4) applies quadratic polynomial in longitude and latitude. Column (5) adds three baseline controls alongside the coordinate polynomial: prior mobilization in 1905, SDP vote share and the log of population in 1916. Column (6) extends the set of controls to include nearest-segment fixed effects, which split the border into ten equally sized segments. In all cases, the coefficient of interest remains similar.

Figure 8 visualizes the reduced form. The figure suggests, that villages just inside Viipuri province had significantly higher war mortality than villages in neighboring provinces in 1918. The estimation in Figure 8, panel (b) corresponds to column (2) of Table 5, panel (c).

As a placebo exercise, Figure A8 shows variation of the reduced form before and after the abolition of censorship. Panel (a) displays before, and panel (b) after. The absence of an effect before the abolition of censorship suggests that Viipuri province was not more mobilized to begin with, but rather turned more mobilized within the ten months after the termination of censorship.

5.3 Further Analysis

Disaggregating War Mortality In the ensuing analysis, I explore different measures of mobilization to understand better, how public opposition is affecting mobilization. First, I split war mortality into different subcategories. Because the change in each subcategory is of no specific interest, I resort to cross-sectional analysis for 1918. Table 3 reports the results.

Columns (1) and (2) show, that the effect of public opposition on war mortality is almost entirely driven by its impact on the rebellious Reds. The coefficient of White war mortality on public opposition in column (2) is small and statistically insignificant. Column (3) reassures this observation, by demonstrating that the share of red casualties was higher in municipalities with a greater increase in public opposition. Columns (1) to (3) thus suggest, that the increase in public opposition merely fostered violence of the rebels.

Similarly, columns (4) and (5) indicate that public opposition was associated with more organized rebel violence. Column (4) shows, that in places with a bigger increase in inflammatory words, a greater proportion of Red casualties belonged to a Red Guard, a paramilitary organization of the rebels. In contrast, column (5) signifies that the increase in public opposition was not associated with a greater proportion of White Guard casualties.

Put together, these findings support the idea that the increase in public opposition achieved to only contribute to the mobilization of the insurgents.

Investigations Another way to assess local mobilization in the civil war is to examine the investigated. Moving forward, I test whether the change in public opposition is associated with the number of post-conflict investigations per capita. I collect data of 6000 investigation records that belong to the Civil War Archives (*Vapaussodan arkisto*) at the National Archives of Finland. Table 3, column (6) verifies a positive association. A one-word increase in exposure to inflammatory rhetoric predicts a 11-log point increase in investigations. The result assures that public opposition was not only related to extreme forms of violence, but can explain participation into the insurgency more broadly.

Paramilitary Organizations To examine the time profile of mobilization, I turn to study the foundation of Red Guards. Using the carefully researched foundation dates for each existing Red Guard in Salkola (1985), I estimate cumulative hazard functions for municipalities with above or below median change in inflammatory words, respectively. I use the Nelson-Aalen estimator:

$$\hat{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i} \tag{7}$$

where d_i is the number of Red Guards founded up to t_i, and n_i is the total municipalities at risk at t_i.

Figure A3 displays the cumulative hazard functions. It shows that municipalities with a larger increase in public opposition were more at risk of setting up a Red Guard in the months leading up to the conflict. Notably, the cumulative hazard functions experience a steep increase and divergence after about 220 days since the foundation of first guards. This period corresponds to late October, 1917.

To get a more concrete picture of the relationship between public opposition and the foundation of Red Guards, I estimate a Cox proportional hazard model:

$$h(t \mid \boldsymbol{x}) = h_0(t)\exp(\boldsymbol{x'\beta})$$

where $\boldsymbol{x}$ is a vector of regressors, including the change in public opposition and all the baseline controls. Table 3, column (7) reports the result. A coefficient of 1.18 implies that an increase in inflammatory rhetoric by one word is associated with a 1.18 times faster foundation of Red Guards, relative to the baseline of zero change.

Evidence from the survival analysis seems to indicate, that public opposition was connected to a faster militarization before the start of the civil war. It also reproduces the well-known observation (Salkola 1985), that the foundation of Red Guards really accelerated around late October of 1917. According to historical research, several trajectories precipitated the surge in militarization simultaneously at the time. One of the most salient factors were the public request for proletariat to organize, composed by Finnish Trade Union Federation and published in left-leaning newspapers, starting in 20$^{\text{th}}$ of October. Another was the strengthening of bolsheviks, and eventually the October Revolution.

Strikes The final perspective from which I explore mobilization is strikes. After being illegalized for three years, striking became a prominent way to protest in 1917. Overall, some 500 strikes were organized, which 150,000 workers participated in. The 1917 demonstrations culminated in a general strike, which was organized a week subsequent to the Bolshevik Revolution, in direct response to Lenin's counsel advocating street protests.

Figure A4 shows event-study estimates from equation (3), while using the log number of strikers per capita as the outcome variable. Akin to mortality, striking shows no evidence of diverging trends by exposure to public opposition before the abolition of censorship. Conversely, the year 1917 witnessed a marked increase in striking activities in regions that were more exposed.

Heterogeneous Effects Having established the main effect of public opposition on mobilization, I now turn to examine whether this effect varies across different contexts. Understanding heterogeneous effects is crucial not only for refining the theoretical framework but also for deriving policy implications related to the conditions under which public dissent is most likely to translate into large-scale mobilization.

To explore potential sources of heterogeneity, I estimate the triple-differences equation (4) by interacting the change in public opposition with key municipal characteristics. Specifically, I consider three dimensions that plausibly shape the extent to which public dissent influences mobilization: (1) the strength of the socialist voter base, (2) economic inequality as measured by land Gini, and (3) the presence of workers' houses that could facilitate rebel organization. These factors capture different mechanisms through which public opposition could be amplified—whether through ideological commitment, economic grievances, or access to organizational infrastructure.

Table A1 presents the estimates of the triple-differences regressions. Across all three specifications, I find that the interaction terms are positive and statistically significant. This suggests that public opposition had a particularly strong mobilizing effect in municipalities with a strong socialist voter base, high levels of economic inequality, and the presence of workers' houses.

First, the results indicate that public opposition had a stronger effect in municipalities where the socialist party had received higher vote shares in prior elections. This is consistent with the notion that ideological alignment with opposition rhetoric plays a crucial role in fostering mobilization. In municipalities where socialism had strong historical support, the increase in pro-opposition rhetoric likely resonated more with local populations, reinforcing prior grievances.[5]

Second, the effect of public opposition was amplified in more economically unequal municipalities, as measured by the land Gini coefficient. This finding aligns with theories of economic grievances as a driver of collective action (Acemoglu and J. Robinson 2006; Meriläinen, Mitrunen, and Virkola 2022). In more unequal regions, the perceived benefits of rebellion were likely higher, making individuals more responsive to shifts in public sentiment.

Finally, municipalities that had workers' houses – local venues used by labor unions and socialist organizations – exhibited a particularly pronounced response to public opposition. The presence of such infrastructure likely lowered the organizational barriers to collective action by providing spaces where dissenters could coordi-

[5] People's predisposition also affected the effectiveness of Nazi indoctrination: see Adena et al. (2015) and Voigtländer and Voth (2015).

nate and plan. This result reinforces the argument that mobilization depends not only on grievances but also on the availability of resources for collective action (Satyanath, Voigtländer, and Voth 2017; Dippel and Heblich 2021).

Taken together, these findings indicate that the effect of public opposition on mobilization was not uniform. Instead, it was conditional on ideological, economic, and organizational factors that shaped the extent to which opposition rhetoric could translate into action.

6 Mechanisms

I now focus on mechanisms behind the effect of newspaper rhetoric on mobilization. I divide the analysis in three parts. First, I study *why* public opposition increased more in some places than others in the first place. I restrict my focus on two related explanations: the abolition of censorship and the example set by the Russian revolution of 1917. Second, I explore *how* public opposition affected the rebels, by utilizing individual-level data on mobilization motives in the interrogation records. I assess whether rebels in municipalities with varying levels of pro-opposition rhetoric were systematically different in terms of their reported motives. Third, I investigate the role of government repression and its interaction with public opposition. Specifically, I examine how state crackdowns shaped mobilization dynamics and whether rising dissent altered the response to repression.

6.1 Spatial Variation of Local Change in Public Opposition

6.1.1 Censorship and Pent-Up Opposition

I hypothesise, that the abolition of censorship unleashed pent-up opposition in the newspapers, and via that increased mobilization. This hypothesis is challenging to test with regional data, because censorship was nationwide policy, and in principle applied to all newspapers. There was varying scrutiny, however, depending on the diligence of the local censor, and how dicey the journal was viewed in the eyes of Russian Imperial authorities (Leino-Kaukiainen 1984). I measure the extent of censorship by journal by computing the fraction of articles read and approved by a censor, and thus carrying a censorship label. To identify municipal variation in censorship, I construct a weighted average of censorship exposure by municipality, so that the weights are based on the subscribership data described in Section 3.

Figure 5 displays the correlation between exposure to censorship and the subsequent change in public opposition. Panel (a) shows the relationship at journal-level, and panel (b) at municipal level. Panel (b) also includes the set of baseline controls. The figure illustrates a strong, positive correlation in both specifications, which are tabulated in columns (1) and (2) of Table 4. Further, column (5) in Table 4 reports the reduced form effect of censorship on war mortality. The estimate suggests that censorship had a positive effect on war mortality, proxying mobilization.

These results are consistent with the hypothesis that the lifting of censorship triggered a surge of bottled-up opposition within the press, which in turn lead to greater mobilization. The findings substantiate previous, qualitative research on the Finnish civil war, which document a major discontinuity in the newspaper rhetoric after the abolition of censorship (Turunen 2021), and suggest that it normalized political violence over the course of 1917 (Ihalainen 2017).

6.1.2 Jumping the Bandwagon

The removal of censorship and succeeding upturn in public opposition in Grand Duchy of Finland are inseparably contingent with the political uproar in Imperial Russia in 1917. February revolution of 1917 effectively ended censorship, and created a power vacuum in Finland which was a fertile environment for an insurgency. Based on suggestive evidence, Russian bolsheviks potentially encouraged, even "deceived" Finnish proletariat to take up arms. One can confidently say that the actions and rhetoric of bolsheviks at minimum inspired Finnish opposition in their own agenda (Salkola 1985; Ihalainen 2017; Turunen 2021).

Figure A5 provides anecdotal evidence of the spread of revolutionary temper from Russia to Finland. The plot overlays the fraction of municipalities which had founded a Red Guard at each date, alongside with trending news terms for selected months. The news terms are identified using TF-IDF, so that they present topics that were particularly emblematic during the month in question. The figure shows, that a big leap in organization of the rebel force coincides with news surge regarding the October revolution. In November 1917, the fraction of municipalities with their own rebel organization shoots from 20% to almost 65%. Meanwhile, the most trending terms in the news consist of words such as "petersburg", "bolshevik", "kordelin", "'krylenko", "people's comissariat" and "military-revolutionary", all describing simultaneous uprising in the neighboring metropole.

Drawing on historical work, my other hypothesis is that public opposition grew more in places which were more exposed to news coverage on the Russian Revolution. I quantify the exposure to news coverage pertaining to the Russian Revolution through the usage of words "bolshevik", "menshevik", "Lenin" and "Trotski". Similar to my main treatment variable, I compute a weighted average of the change in revolution news coverage by municipality, using the municipal breakdown of subscribership as weights.

Figure 6 shows the relationship between news coverage on the Russian Revolution and the change in public opposition. Panel (a) shows the relationship at journal-level, and panel (b) at municipal level, while panel (b) also includes the set of baseline controls. Journals that wrote more news regarding the Russian Revolution also began to use more inflammatory language after the removal of censorship. This relationship also applies at municipal level: regions that were more exposed to news coverage on the Russian Revolution experienced a greater increase in public opposition. Thirdly, the more exposed regions witnessed a higher death rate in the civil war. The estimates from these regressions are tabulated in columns (3), (4) and (6) of Table 4.

Russian Telegraph Why did some newspapers write more about the Russian Revolution than others? There are at least three possible explanations: sheer randomness, the editors' political views or connections to Russia. I will focus on the last. It is justifiable to deduce that if some journal editors received a stronger signal of the frailty of the Empire, they felt more safe to publish inflammatory news than the others. Better connections to Russia also provided a better platform for Russian Bolsheviks to spread their revolutionary program.

I test this conjecture by regressing the news coverage on the Russian Revolution on an indicator variable denoting the location of Russian telegraph stations. Telegraph was an irreplaceable tool of communication for newspapers, albeit expensive (Tommila, Nygård, and Salokangas 1987). The Russian telegraph line was built in the 1850s with the upkeep of Imperial business in mind (Suomen maantieteellinen seura 1899). Hence, its location should be orthogonal to any other factors that might affect the mobilization potential of municipalities. Columns (1) and (2) in Table A2 show that municipalities which were stationed with a Russian telegraph office were exposed to more news pertaining the Russian Revolution, and a bigger change in public opposition. Column (3) further shows, that telegraph had a positive effect on war mortality. These findings lend support to the idea that the journals whose editors had closer ties to Russia either felt more safe or persuaded to publish more revolutionary news.

6.2 Mobilization Motives

The motivations behind individual participation in rebellion have long been a focus of political science, with numerous scholars offering insights into what drives people to rise up. In his seminal work, *Why Men Rebel*, Ted Gurr emphasizes the ideological foundations of revolutionary movements. He argues that the primary driver of rebellion is the gap between individuals' rising expectations and their actual capabilities. These heightened expectations, often fueled by ideologies challenging the status quo, have historically sparked uprisings (Gurr 1970).

Another branch of research has explored how individuals may be compelled toward political violence as a result of social pressure. (Bernheim 1994; Bursztyn, Egorov, et al. 2023). Durlauf (2004) highlights, that people have been found to have a psychological tendency to conform to the behavior of others. In the context of the Rwandan genocide, Straus (2007), drawing on interviews with 200 convicted perpetrators, provides anecdotal evidence of "violence begetting violence" (Yanagizawa-Drott 2014). Many perpetrators reported that face-to-face mobilization and fear of sanctions from their own social group compelled them to commit acts of violence.

In addition to intangible incentives, material motives play a significant role in conflict research. Collier and Hoeffler (2004) and Miguel, Satyanath, and Sergenti (2004) have shown that negative income shocks may encourage revolting, when the expected income as a rebel might suddenly exceed the expected income under the status quo. Lack of opportunities during economic downturns with increasing unemployment thus create potential insurgents. Sometimes rebels are enticed by future rewards rather than by present grievances. For instance, the poor might take the streets in order to displace the incumbent elite to induce redistribution of resources (Acemoglu and J. Robinson 2006; Meriläinen, Mitrunen, and Virkola 2022).

Word Cloud To get a first glance of motivations behind the rebellion that started the Finnish Civil War, Figure A11 presents a word cloud of twenty of the most common words describing mobilization motives over the rebels' interrogations. The words are drawn from 50,000 responses to the question "Why did you join the Red Guards, what salary was promised and received?". Stopwords, words regarding the salary and terms constituting the question itself have been filtered out.

From the word cloud, three types of words stand out: terms related to material, ideological and conformity motives. I have differentiated words in these three separate categories by coloring them with yellow, grey and blue, respectively. Material category consists of words mainly related to unemployment, but also the shortage of food. Ideological category includes words indicating voluntary participation, and also those describing joining for undefined reasons. Conformity category amounts of expressions such as "coercion", "had to" and "everyone", i.e. words depicting various type of social pressure, whether it is merely emulating the behaviour of others or at the other extreme being threatened to join the cause.

Text Classification For a more structured understanding of the mobilization motives, I use text classification to categorize each individual motive into one of the aforementioned three classes. Specifically, I estimate a single multiclass text classification model using a simple feed-forward neural network. The three classes were chosen based on the author's own reading of the training data and the previous descriptive findings. Moreover, the selected classes align well with earlier research that has examined samples of the interrogation records (Arosalo 1969; Kekkonen 1991; Arosalo 1998). To get a better idea of what each class of motives is capturing, Figure A12 displays the ten most frequent words by class in the training data.

Next, I proceed to describe the training process. First, I transform the textual data into TF-IDF representations. Once the data is vectorized, I split it into training (80%) and validation (20%) sets. The neural network itself is straightforward: it has two hidden layers (with 128 and 64 units, respectively), both using ReLU activations, followed by a final softmax output layer to produce the class probabilities. I compile the model with the Adam optimizer and use categorical cross-entropy as the loss function. To mitigate overfitting, an early stopping criterion is employed: training halts if the validation loss does not improve for five consecutive epochs, and the best model weights up to that point are automatically restored. After training, the model achieves an accuracy of 82% on the validation set and an F1-score of 0.81, indicating strong overall performance in correctly classifying the motives.

Figure A13 illustrates the fraction of motives classified to contain elements of conformity, ideology or material reasons. By far, the most common reported motive behind revolting was conformity. According to the text classification model, over 40% of the rebels appealed to social pressure to rationalize their participation, typically referring to encouragement or even coercion by their peers. A sizable proportion of the interrogated, 19%, also emphasized economic hardship. Ideology was the least frequent of the three categories, yet still almost 12% of the defendants admitted being driven by sympathy to the cause. For a more tangible perception of the kind of responses in each class, Table A3 presents a random sample of top-scoring motives by class.

Public Opposition and Mobilization Motives How did the change in public opposition affect people's self-perceived motivations to rebel? To answer this question, I estimate a linear probability model

$$y_i = \alpha + \beta \Delta PublicOpposition_m + \boldsymbol{x}_i' \boldsymbol{\gamma} + \varepsilon_i \tag{8}$$

where y_i is and indicator variable, which gets value one for individual i if their motive was classified to belong to a given motive category.

Table A4 reports the results. Columns (1) to (4) suggest, that an increase in pro-opposition rhetoric had small to no effect on the likelihood of reporting any particular motive. That said, there is a minor, statistically significant effect on being motivated by material grievance. Specifically, a one standard deviation increase in inflammatory words increases the probability of appealing to economic hardship by about one percentage point. Relative to the mean of 19%, however, the observed effect is marginal.

The virtual absence of an effect of public opposition on mobilization motives suggests that, at the individual level, the distribution of reasons for rebellion remains stable across different levels of the treatment. This implies that changes in public opposition do not redistribute individuals between the different motivations for joining the uprising. One interpretation is, that the motivating factors that push an individual to join an insurgency for any reason might be deeply personal or situational, and not directly tied to the broader changes in public sentiment.

While inflammatory rhetoric has no observable influence on individual probabilities of different motives, there might be some underlying group-level dynamics that do not manifest clearly at the individual level but become apparent when viewed in the aggregate. To study this conjecture, I re-estimate equation (8) using municipal data. Table A5 views the results.

Columns (1) through (4) reveal, that public opposition has a sizable and statistically significant effect on the *sum* of differently motivated rebels in each motive category. In fact, the effect size is almost identical for ideological, conformity and material rebels, and for those insurgents whose motive could not be classified. These coefficients are qualitatively similar to the baseline estimates in Table 2.

The uniform effect of pro-opposition rhetoric across rebel groups suggests that rising dissent increases overall participation without altering the distribution of motivations, as shown in Figure A15. As public opposition grows, all factions expand proportionally, implying that participation is driven by a shared catalyst rather than shifts in individual motives.

The political economy literature on global games postulates, that this pattern may emerge if public opposition beholds strategic information about the beliefs of other people (Carlsson and Van Damme 1993; Morris and Shin 2001; Vives 2005; Angeletos, Hellwig, and Pavan 2007; Edmond 2013). Specifically, rising dissent may signal that the masses are turning against the regime, prompting individuals to update their beliefs about the success of rebellion. Public opposition thus functions through *strategic complementarity*: as visible dissent increases, individuals are more likely to join, believing broader collective action raises the probability of success. This dynamic can motivate participation across all types of rebels.

Public Opposition and Strategic Complementarity Do rebels' decisions exhibit strategic complementarity? To examine this, I need to know the timing of mobilization for each rebel, as well as their peers, and see whether the two move in parallel. To this end, I use the interrogation records to construct a variable denoting the month in which each rebel got enlisted to a Red Guard. The timing of enlistment is derived from responses to the question "When did you join the Red Guard: which regiment, batallion, company, section, (platoon); names of commanders, staff members and other officers of the guard?". Due to the question's open-ended nature, I construct the enlistment date variable in three steps: first, I use NER to recognize any dates in the responses. Second, I vectorize any date-candidates by employing a Sentence transformer. This step is crucial in dealing with typographical errors, inherited from the HTR. Third, because the date-candidates come in heterogeneous formats, I manually classify a subset of 850 dates, and then train a feed-forward neural net to identify the remaining enlistment dates, using the sentence embeddings as inputs. The resulting classifier is highly robust, being able to classify date-candidates with 97% accuracy in validation data.

To investigate whether rebels' decision to revolt exhibits strategic complementarity, I estimate a discrete-time hazard model (Jenkins 1995; Nunn and Qian 2014; Moscona, Nunn, and J. A. Robinson 2020)

$$P(T_{im} = t \mid T_{im} \geq t; \boldsymbol{x}_{mt}) = F(\alpha_m + \theta(t) + \beta PeerMobilization_{mt}), \tag{9}$$

where t indexes time and T_{im} is the date of enlistment for individual i in municipality m. $PeerMobilization_{mt}$ is the local leave-one-out mobilization rate in municipality m at month t. The sample includes all individuals that are at risk for mobilizing, or all rebels that have not joined a Red Guard thus far. The dependent variable is the discrete-time hazard rate $h_{imt} = P(T_{im} = t \mid T_{im} \geq t; \boldsymbol{x}_{mt})$, while F is a logistic CDF, standard normal CDF or an identity function, depending on specification. α_m is a municipality fixed effect, while $\theta(t)$ is estimated by a fifth-order polynomial or a full set of month fixed effects. The coefficient of interest is β, which measures the extent of strategic complementarity, i.e. how strongly individuals' decision to revolt depends on the actions of their peers.

The estimates from equation (9) are reported in Table A6. The baseline model in column (1) assumes that the hazard rate h_{it} follows a logistic distribution, and estimates $\theta(t)$ as a polynomial. Column (2) is otherwise identical, but assumes a Gaussian distribution. Column (3) employs identity function as the hazard function, while estimating $\theta(t)$ flexibly with month fixed effects. These estimates express, that individual's resolution to take the streets is correlated with the action their peers take. Specifically, increasing the local mobilization rate from zero to one percent – which is close to the yearly sample average – is associated with a 5 p.p. increase in the individual mobilization probability, on average. According to the logistic discrete-time hazard model in column (1), this corresponds to a $(\exp(0.98) - 1) \cdot 100 = 166\%$ increase in odds. Figure A16 of predicted hazards illustrates, what this means in practice. For instance, in February 1918, when mobilization was at its fastest, the median predicted hazard rate was around 50% in municipalities with peer mobilization rate of 1%, in comparison to 25% in municipalities with total abstention.

Does public opposition beget mobilization via strategic complementarity? If so, we would expect that dissent has a larger effect when it becomes common knowledge. Morris and Shin (2001) name this phenomenon as "publicity multiplier": if every individual knows that they are receiving the same information as their fellow citizens, and are also concerned about their action, then publicity will reinforce the impact of public signals beyond their information content. I examine the publicity effect by testing whether the mobilizing impact of dissent is aggravated when (1) it reaches a wider audience, or (2) the signal of discontent is more overt. To explore these hypotheses, I estimate the triple-difference equation (4), focusing on the differential effects of two variables in particular: the number of newspaper subscribers per capita and the number of strikers per capita.

Columns (4) and (5) of Table A6 summarize the results. Column (4) shows, that public opposition had a stronger effect on war mortality in municipalities with above median number of newspaper subscribers per capita. Column (5) demonstrates, that this was also the case in regions with above median number of strikers per person in the year 1917. These results provide evidence supporting the interpretation that public opposition affect people's mobilization decision mainly by shifting their beliefs about others' intentions. Appendix E outlines a simple global game model consistent with this interpretation.

Limitations Drawing on interrogation records as a data source entails certain limitations. The most troubling issue is that the accused had an incentive to lie. According to Kekkonen (1991), it is plausible that the defendants sought to appeal on motives they believed would mitigate their culpability. During interrogations, the accused gave information at the risk of their own life and freedom, which could make the data quality questionable. What is more, some motives, such as compulsion, were considered as mitigating factors by the courts. Note, however, that the courts were fully aware of the accused's temptation to deviate from truth-telling. To take this into account, false testimony was disincentivized in turn by making it an aggravating issue when deciding on sentences. Still, presumably some captives thought it safer to whitewash their part in the rebellion. To validate the informational content of the accused's motives, I examine how they correlate with key covariates in a linear probability model.

Figure A14 presents the results. The plot shows that rebels in smaller municipalities were more likely to join the Red Guard due to peer pressure, while those from areas with strong SDP support more often cited ideological motives. In contrast, rebels in larger municipalities were more likely to appeal to material reasons. These patterns are intuitive: social pressure was likely stronger in small rural communities, reinforcing incentives to conform. In municipalities with a strong socialist voter base, ideological alignment with the rebellion was more prevalent. Lastly, since unemployment was particularly high in large towns, it is unsurprising that economic hardship played a greater role in motivating rebels there.

6.3 Repression

One important aspect of mobilization that we have thus far neglected is repression. When facing an existential threat, autocrats often resort to the costly strategy of repressing the opposition. Although intended to suppress dissent, repression may have an ambiguous effect on mobilization. On the one hand, successful repression can directly stifle revolutionary action by deterring participation. On the other hand, it may delegitimize the regime, intensify grievances, and trigger backlash mobilization.

I examine the relationship between repression and mobilization by estimating an event-study regression with staggered treatment (Callaway and Sant'Anna 2021; Roth et al. 2023). In this design, the average treatment effect on the treated (ATT) in period t for municipalities repressed from period g onward is given by

$$ATT(g,t) = E[Y_{m,t} - Y_{m,g-1} \mid G_m = g] - E[Y_{m,t} - Y_{m,g-1} \mid G_m = g'] \quad \text{for any } g' > t,$$

where $Y_{m,t}$ is the log of the mobilization rate in municipality m at month t. The average ATT – or event-study parameter – l periods after repression is

$$ATT_l^w = \sum_g w_g ATT(g, g+l). \tag{10}$$

I measure the onset of repression in two ways: first, by the date of the first local rebel casualty, and second, by the date of the first major battle, both drawn from WarVictimSampo.

Estimates of (10) are plotted in Figure A17. I find compelling evidence that the onset of repression is followed by declining mobilization. This pattern is confirmed by two-way fixed effects regressions, reported in columns (1) and (3) of Table A7.

Next, I examine whether the effect of repression varies by the level of public dissent. Theoretically, repression could have a differential impact depending on the extent of opposition already present in a municipality. If repression is more costly for autocrats in areas with high public dissent – due to the risk of radicalizing further segments of the population – it could mitigate the demobilizing effect. Conversely, repression could be particularly effective in such areas if dissenters who were drawn in by the momentum are not as committed to the cause.

To test this hypothesis, I estimate a modified version of the TWFE regression that allows the effect of repression to differ across municipalities with varying levels of public dissent. The estimates, reported in Table A7, suggest that repression is associated with a weaker decline in mobilization in municipalities with above median change in public dissent. However, this difference is not statistically significant. Thus, while repression may not be equally effective across all contexts, I find no conclusive evidence that dissent meaningfully moderates its impact.

Overall, these results reinforce the argument of that repression is a powerful tool for neutralizing collective action. While public dissent provides a signal of potential mobilization, it does not offer immunity against oppression.

7 Robustness

7.1 Matching on Preexposure Characteristics

A related concern to demand-based selection is, that the control variables fail to form a plausible counterfactual region group for more exposed municipalities. To mitigate such worry, I re-estimate the difference-in-differences specification after matching municipalities on preexposure characteristics. I use baseline controls as the matching variables, and employ two different matching algorithms: nearest neighbor propensity score matching and propensity score subclassification. Table A8 shows, that the results are qualitatively comparable to the baseline.

7.2 Alternative Specifications

I continue to assess, whether the results hinge on assumptions regarding the variance covariance matrix, outlier observations or definition of the exposure variable. Table A9 suggests that this is not the case. In column (1), I show that the baseline estimate is robust to using Conley Spatial HAC standard errors, while adjusting for spatial correlation within 100 kilometer radius. In column (2), I omit Viipuri province from the sample, a region where the change in public opposition was particularly drastic. The coefficient of interest remains in the ballpark of the baseline estimate.

In column (3), I reconstruct the treatment variable, while using an alternative, empirical proxy of local journal composition. Concretely, I conduct named-entity recognition for the universe of newspaper articles before the abolition of censorship, and measure subscribership for each journal by how often the journal mentions a given municipality. The idea behind the proxy is, that local newspapers are expected to write news about local events, and even national newspapers probably give more coverage to stories in their most important market areas. Thus, municipality references could serve as a rough indicator for subscribers. An important advantage of determining empirical journal composition is, that it does not suffer from data limitations, allowing me to exploit 100% of newspaper circulation when measuring the municipal change in public opposition. Despite the fact that the proxy exposure is probably subject to nontrivial measurement error, the resulting estimate is roughly comparable to the baseline.

7.3 Additional Controls

Table A10 shows, that the baseline result is robust to the inclusion of additional control variables. One potential concern is, that my baseline controls for the vote share of social democratic party and prior mobilization in 1905 are unable to capture regional differences in radicalism. In column (1), I further control for the total number of strikes in 1910–1914, and the number of workers' houses in 1916. This addition does not effect the estimates.

Previous literature has found, that deprivation, ethnic diversity, terrain ruggedness and inequality are important drivers of political violence (Fearon and Laitin 2003; Miguel, Satyanath, and Sergenti 2004; Montalvo and Reynal-Querol 2005;

Meriläinen, Mitrunen, and Virkola 2022). Columns (2) through (5) of Table A10 demonstrate, that my result is robust to controlling for the severity of food shortage in 1917 (Rantatupa 1979), the share of Swedish-speaking population within municipality, the log of terrain slope and land Gini in 1910 (Meriläinen, Mitrunen, and Virkola 2022).

According to some historical accounts, the presence Russian soldiers was an important impetus in launching the atrocities in 1918 (Rasila 1969). In column (6), I verify that my results are not driven by the stationing of Russians, by controlling for the log total of Russian troops by municipality.

In column (7), I further account for the possibility of unobserved regional trends by including province-specific slopes. This adjustment allows for differential trends in public opposition across provinces, capturing potential variations in underlying political dynamics. The estimates remain robust, suggesting that the baseline results are not driven by regional differences in the rate of change of opposition or mobilization.

8 Conclusion

This paper examines how a shift in collective sentiment, triggered by the 1917 Russian Revolution, led to mass mobilization in Finland. I show how the removal of imperial censorship, coupled with the diffusion of revolutionary ideas from Petrograd, generated a sharp rise in public opposition, reflected in newspaper rhetoric. As dissent became more visible and expectations about the feasibility of rebellion shifted, mobilization accelerated. Ten months later, Finland descended into civil war. I argue that this surge in opposition was driven by two interrelated forces: the release of long-suppressed grievances and the demonstration effect of the Bolshevik Revolution.

I supplement these findings by analyzing interrogation records, which document the rebels' self-attributed motivations for mobilization. I find that public opposition fueled rebellion primarily by shifting beliefs about the likelihood of success rather than altering individuals' intrinsic motives. Dissent served as a rallying cry, increasing participation without changing the underlying reasons for rebellion. Mobilization was also highly responsive to peer influence—41% of rebels cited encouragement, coercion, or communal expectations as key factors in their decision to take up arms. However, government repression had a strong demobilizing effect, which public opposition failed to counteract. This suggests that while belief shifts can trigger rebellion, they are possibly insufficient to sustain it in the face of state coercion.

The results in this article provide broader lessons on the role of popular sentiment in revolutionary movements. The sequence of events in 1917 Finland illustrate, that the legitimacy of an incumbent government can erode fast, and is particularly vulnerable to signals of successful uprisings elsewhere. This outcome is exceptionally indicative considering that it was unexpected of Finland, perceived as a “conservative” and “counter-revolutionary” country (Hobsbawm 1995). However, factors such as fear, social pressure or pluralistic ignorance can explain why seemingly stable societies may under given circumstances fall like a house of cards (Kuran 1989; Bursztyn, González, and Yanagizawa-Drott 2020). Similar dynamics have played out in more recent revolutionary waves, from the uprisings of 1989 (Huntington 1993; Lohmann 1994) to the color revolutions of the early 2000s (Strömberg 2015) and the Arab Spring in 2011 (Tilly and Wood 2015), where governments lost control of the dominant narrative, allowing opposition to mobilize at scale. A compelling revolutionary message can unite dissenters and destabilize regimes before they can mount an effective response (Goldstone 2023).

Yet modern autocrats have adapted to these risks, recognizing that controlling the public narrative is often more effective than outright repression (Guriev and Treisman 2022). Instead of relying solely on coercion, they manipulate perceptions to maintain an illusion of mass support and discourage collective action. One strategy is the use of state-controlled media and online disinformation campaigns, such as troll factories, which flood public discourse with regime-aligned content to fabricate the appearance of widespread approval (Stukal, Sanovich, Bonneau, et al. 2017; Stukal, Sanovich, Tucker, et al. 2019). By shaping what people believe about the popularity of dissent, these tactics exploit the same social mechanisms that fuel revolutions.

References

Acemoglu, Daron, Naidu, Suresh, et al. (2019). "Democracy does cause growth." In: *Journal of political economy* 127.1, pp. 47–100.

Acemoglu, Daron and Robinson, James (2006). *Economic origins of dictatorship and democracy.* Cambridge University Press.

Adena, Maja et al. (2015). "Radio and the Rise of the Nazis in Prewar Germany." In: *The Quarterly Journal of Economics* 130.4, pp. 1885–1939.

Aidt, Toke S and Jensen, Peter S (2014). "Workers of the world, unite! Franchise extensions and the threat of revolution in Europe, 1820–1938." In: *European Economic Review* 72, pp. 52–75.

Alapuro, Risto (2018). *State and revolution in Finland.* Brill.

Aminoff, Gustaf (1918). *Kapinaliikkeen alkuvalmistelut, sen kehitys ja leviäminen tammikuun 31 päivään 1918.* Helsinki: Suomen Senaatin kirjapaino.

Angeletos, George-Marios, Hellwig, Christian, and Pavan, Alessandro (2007). "Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks." In: *Econometrica* 75.3, pp. 711–756.

Arosalo, Sirkka (1969). *"Punaisten" tuomiot tuomiot vuonna 1918: Tampereen piirin valtiorikosoikeuksissa annettujen tuomioitten määräämisperusteet kuulustelupöytäkirjojen ja suojeluskuntien lausuntojen valossa.* Tampere.

— (1998). "Social conditions for political violence: red and white terror in the Finnish Civil War of 1918." In: *Journal of Peace Research* 35.2, pp. 147–166.

Ash, Elliott and Hansen, Stephen (2023). "Text algorithms in economics." In: *Annual Review of Economics* 15.1, pp. 659–688.

Åström, L. (1918). "Sanomalehtien kautta harjoitetusta joukkosuggestionista. Vapaan sanan vallan laajentaminen. Yhteiskunnallis-sielutieteellinen tutkielma." In: *Aika*, pp. 7–25.

Bai, Ying, Jia, Ruixue, and Wang, Runnan (2023). *Ideas Mobilize People: The Diffusion of Communist Ideology in China.* Tech. rep. National Bureau of Economic Research.

Baker, Scott R, Bloom, Nicholas, and Davis, Steven J (2016). "Measuring economic policy uncertainty." In: *The quarterly journal of economics* 131.4, pp. 1593–1636.

Bernheim, B Douglas (1994). "A theory of conformity." In: *Journal of political Economy* 102.5, pp. 841–877.

Boillet, Mélodie, Kermorvant, Christopher, and Paquet, Thierry (2021). "Multiple document datasets pre-training improves text line detection with deep neural networks." In: *2020 25th International conference on pattern recognition (ICPR).* IEEE, pp. 2134–2141.

Boillet, Mélodie, Kermorvant, Christopher, and Paquet, Thierry (2022). "Robust text line detection in historical documents: learning and evaluation

methods." In: *International Journal on Document Analysis and Recognition (IJDAR)* 25.2, pp. 95–114.
Bursztyn, Leonardo, Egorov, Georgy, et al. (2023). "Justifying dissent." In: *The Quarterly Journal of Economics* 138.3, pp. 1403–1451.
Bursztyn, Leonardo, González, Alessandra L, and Yanagizawa-Drott, David (2020). "Misperceived social norms: Women working outside the home in Saudi Arabia." In: *American economic review* 110.10, pp. 2997–3029.
Callaway, Brantly and Sant'Anna, Pedro HC (2021). "Difference-in-differences with multiple time periods." In: *Journal of econometrics* 225.2, pp. 200–230.
Carlsson, Hans and Van Damme, Eric (1993). "Global games and equilibrium selection." In: *Econometrica: Journal of the Econometric Society*, pp. 989–1018.
Collier, Paul and Hoeffler, Anke (2004). "Greed and grievance in civil war." In: *Oxford economic papers* 56.4, pp. 563–595.
Conley, Timothy G (1999). "GMM estimation with cross sectional dependence." In: *Journal of econometrics* 92.1, pp. 1–45.
Dahl, Christian M et al. (2023). "Applications of machine learning in tabular document digitisation." In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 56.1, pp. 34–48.
De Mesquita, Ethan Bueno (2010). "Regime change and revolutionary entrepreneurs." In: *American Political Science Review* 104.3, pp. 446–466.
Dippel, Christian and Heblich, Stephan (2021). "Leadership in social movements: Evidence from the "forty-eighters" in the civil war." In: *American Economic Review* 111.2, pp. 472–505.
Durlauf, Steven N (2004). "Neighborhood effects." In: *Handbook of regional and urban economics* 4, pp. 2173–2242.
Edmond, Chris (2013). "Information manipulation, coordination, and regime change." In: *Review of Economic studies* 80.4, pp. 1422–1458.
Eichengreen, Barry J (1992). *Golden fetters: the gold standard and the Great Depression, 1919-1939.* NBER series on long-term factors in economic development.
Enikolopov, Ruben, Makarin, Alexey, and Petrova, Maria (2020). "Social media and protest participation: Evidence from Russia." In: *Econometrica* 88.4, pp. 1479–1514.
Enikolopov, Ruben, Petrova, Maria, and Zhuravskaya, Ekaterina (2011). "Media and political persuasion: Evidence from Russia." In: *American Economic Review* 101.7, pp. 3253–3285.
Enke, Benjamin (2020). "Moral values and voting." In: *Journal of Political Economy* 128.10, pp. 3679–3729.
Fearon, James D and Laitin, David D (2003). "Ethnicity, insurgency, and civil war." In: *American political science review* 97.1, pp. 75–90.

Gennaro, Gloria and Ash, Elliott (2022). "Emotion and reason in political language." In: *The Economic Journal* 132.643, pp. 1037–1059.

Gentzkow, Matthew, Kelly, Bryan, and Taddy, Matt (2019). "Text as data." In: *Journal of Economic Literature* 57.3, pp. 535–574.

Goldsmith-Pinkham, Paul, Sorkin, Isaac, and Swift, Henry (2020). "Bartik instruments: What, when, why, and how." In: *American Economic Review* 110.8, pp. 2586–2624.

Goldstone, Jack A (2023). *Revolutions: A very short introduction.* Vol. 381. Oxford University Press.

Guriev, Sergei, Melnikov, Nikita, and Zhuravskaya, Ekaterina (2021). "3g internet and confidence in government." In: *The Quarterly Journal of Economics* 136.4, pp. 2533–2613.

Guriev, Sergei and Treisman, Daniel (2022). *Spin dictators: The changing face of tyranny in the 21st century.* Princeton University Press.

Gurr, Ted Robert (1970). *Why men rebel.* Routledge.

Haapala, Pertti (1995). *Kun yhteiskunta hajosi: Suomi 1914-1920.* Painatuskeskus.

Hannula, Joose Olavi and Korhonen, Arvi (1956). *Suomen vapaussodan historia.* WSOY.

Hobsbawm, Eric (1995). *Age of extremes: the short twentieth century, 1914-1991.* Abacus London.

Horton, John J (2023). *Large language models as simulated economic agents: What can we learn from homo silicus?* Tech. rep. National Bureau of Economic Research.

Huntington, Samuel P (1993). *The third wave: Democratization in the late twentieth century.* Vol. 4. University of Oklahoma press.

Ihalainen, Pasi (2017). *The Springs of democracy: National and transnational debates on constitutional reform in the British, German, Swedish and Finnish parliaments, 1917-1919.* Finnish Literature Society/SKS.

Jenkins, Stephen P (1995). "Easy estimation methods for discrete-time duration models." In: *Oxford Bulletin of Economics & Statistics* 57.1.

Kekkonen, Jukka (1991). "Laillisuuden haaksirikko: rikosoikeudenkäyttö Suomessa vuonna 1918." In.

Kirby, David G (1978). "Revolutionary ferment in Finland and the origins of the Civil War 1917–1918." In: *Scandinavian Economic History Review* 26.1, pp. 15–35.

Korinek, Anton (2023). "Generative AI for economic research: Use cases and implications for economists." In: *Journal of Economic Literature* 61.4, pp. 1281–1317.

Kuran, Timur (1989). "Sparks and prairie fires: A theory of unanticipated political revolution." In: *Public choice*, pp. 41–74.

— (1991). “The East European revolution of 1989: is it surprising that we were surprised?” In: *The American Economic Review* 81.2, pp. 121–125.

Kuusanmäki, Jussi (1980). “Ensimmäisen maailmansodan lehdistösensuuri.” In: *Sensuuri ja sananvapaus Suomessa*. Ed. by Pirkko Leino-Kaukiainen. Suomen Sanomalehdistön Historia -projektin julkaisuja 17. Helsinki, pp. 85–105.

Leino-Kaukiainen, Pirkko (1984). *Sensuuri ja sanomalehdistö Suomessa vuosina 1891-1905= Censorship and newspapers in Finland, 1891-1905*. Suomen historiallinen seura.

Li, Minghao et al. (2023). “Trocr: Transformer-based optical character recognition with pre-trained models.” In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. 11, pp. 13094–13102.

Lindfors, Teppo, Meriläinen, Jaakko, and Mitrunen, Matti (2024). “Kriisit ja yhteiskunnallinen muutos – luotiinko sisällissodan vuosina tasa-arvoisemman yhteiskunnan edellytykset?” In: *Ankarat ajat — Suomalaisten talouskriisien pitkä historia*. Ed. by Jari Eloranta and Roope Uusitalo.

Lippmann, Quentin (2022). “Gender and lawmaking in times of quotas.” In: *Journal of Public Economics* 207, p. 104610.

Lohmann, Susanne (1994). “The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989–91.” In: *World politics* 47.1, pp. 42–101.

Luoma, Jouni et al. (May 2020). “A Broad-coverage Corpus for Finnish Named Entity Recognition.” In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Ed. by Nicoletta Calzolari et al. Marseille, France: European Language Resources Association, pp. 4615–4624. URL: https://aclanthology.org/2020.lrec-1.567.

Manacorda, Marco and Tesei, Andrea (2020). “Liberation technology: Mobile phones and political mobilization in Africa.” In: *Econometrica* 88.2, pp. 533–567.

Matikainen, Juha (2018). “Parlamentarismin kannattajasta vallankumouksen äänitorveksi: Suomen Sosialidemokraattisen Puolueen lehdistö 1917-1918.” In: *Jyväskylä studies in humanities* 338.

Meriläinen, Jaakko, Mitrunen, Matti, and Virkola, Tuomo (2022). “Famine, Inequality, and Conflict.” In: *Journal of the European Economic Association*, jvac069.

Miguel, Edward, Satyanath, Shanker, and Sergenti, Ernest (2004). “Economic shocks and civil conflict: An instrumental variables approach.” In: *Journal of political Economy* 112.4, pp. 725–753.

Mikolov, Tomas et al. (2013). “Distributed representations of words and phrases and their compositionality.” In: *Advances in neural information processing systems* 26.

Montalvo, José G and Reynal-Querol, Marta (2005). "Ethnic polarization, potential conflict, and civil wars." In: *American economic review* 95.3, pp. 796–816.

Morris, Stephen and Shin, Hyun Song (2001). "Global games: Theory and applications." In.

Moscona, Jacob, Nunn, Nathan, and Robinson, James A (2020). "Segmentary Lineage Organization and Conflict in Sub-Saharan Africa." In: *Econometrica* 88.5, pp. 1999–2036.

Nunn, Nathan and Qian, Nancy (2014). "US food aid and civil conflict." In: *American economic review* 104.6, pp. 1630–1666.

Qi, Peng et al. (2020). "Stanza: A Python natural language processing toolkit for many human languages." In: *arXiv preprint arXiv:2003.07082*.

Rantatupa, Heikki (1979). *Elintarvikehuolto ja-säännöstely Suomessa vuosina 1914-1921*. 17. Jyväskylän yliopisto.

Rasila, Viljo (1969). "The Finnish civil war and land lease problems." In: *Scandinavian Economic History Review* 17.1, pp. 115–135.

Reed, John (1919). *Ten days that shook the world*. Haymarket Books.

Reimers, Nils and Gurevych, Iryna (Nov. 2019). "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. URL: http://arxiv.org/abs/1908.10084.

— (Nov. 2020). "Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation." In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. URL: https://arxiv.org/abs/2004.09813.

Roth, Jonathan et al. (2023). "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." In: *Journal of Econometrics* 235.2, pp. 2218–2244.

Ryömä, Hannes (1918). *Vallankumousvuoden tapahtumista*. Uusi kiriapaino.

Salkola, Marja-Leena (1985). *Työväenkaartien synty ja kehitys punakaartiksi 1917-18 ennen kansalaissotaa*. Vol. 1. Valtion painatuskeskus.

Santavirta, Torsten and Stuhler, Jan (2020). "Name-based estimators of intergenerational mobility." In: *Technical Report*.

Satyanath, Shanker, Voigtländer, Nico, and Voth, Hans-Joachim (2017). "Bowling for fascism: Social capital and the rise of the Nazi Party." In: *Journal of Political Economy* 125.2, pp. 478–526.

Shen, Zejiang et al. (2021). "Layoutparser: A unified toolkit for deep learning based document image analysis." In: *Document Analysis and Recognition–ICDAR 2021: 16th International Conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part I 16*. Springer, pp. 131–146.

Siltala, Juha (2023). “Cumulative frustrations and the erosion of social trust before the Finnish Civil War.” In: *Scandinavian Journal of History* 48.2, pp. 157–178.

Snyder Jr, James M and Strömberg, David (2010). “Press coverage and political accountability.” In: *Journal of political Economy* 118.2, pp. 355–408.

Stenhammar, Aapo (2023). “Tensions of Structural Change: Evidence from Finland’s Field Reservation Policy.” Job Market Paper.

Straus, Scott (2007). “What is the relationship between hate radio and violence? Rethinking Rwanda’s “Radio Machete”.” In: *Politics & Society* 35.4, pp. 609–637.

Strömberg, David (2015). “Media and politics.” In: *Annual Review of Economics* 7.1, pp. 173–205.

Stukal, Denis, Sanovich, Sergey, Bonneau, Richard, et al. (2017). “Detecting bots on Russian political Twitter.” In: *Big data* 5.4, pp. 310–324.

Stukal, Denis, Sanovich, Sergey, Tucker, Joshua A, et al. (2019). “For whom the bot tolls: a neural networks approach to measuring political orientation of Twitter bots in Russia.” In: *Sage Open* 9.2, p. 2158244019827715.

Suodenjoki, Sami and Turunen, Risto (2017). *Työväki kumouksessa.* Työväen historian ja perinteen tutkimuksen seura.

Suomen maantieteellinen seura (1899). *Suomen kartasto. Teksti.* Helsinki: Osakeyhtiö F. Tilgmannin kirja- ja kivipaino.

Tikka, Marko (2018). “Hitaasti koottu palapeli: Suomen sisällissota historiankirjoituksessa.” In: *Historiallinen aikakauskirja* 116.2, pp. 118–125.

Tilly, Charles and Wood, Lesley J (2015). *Social Movements, 1768-2012.* Routledge.

Tommila, Päiviö, Nygård, Toivo., and Salokangas, Raimo. (1987). *Suomen lehdistön historia. 2, Sanomalehdistö suurlakosta talvisotaan / päätoimittaja: Päiviö Tommila ; kirjoittajat: Toivo Nygård, Raimo Salokangas.* fin. Kuopio: Kustannuskiila. ISBN: 951-657-198-0.

Tommila, Päiviö, Raitio, Tuire, and Aalto, Eeva-Liisa (1977). *Suomen sanomalehdistön levikki 1861-1900.* Suomen sanomalehdistön historia-projekti.

Turunen, Risto (2021). *Shades of Red.* Työväen historian ja perinteen tutkimuksen seura.

Upton, Anthony F (1980). *The Finnish Revolution: 1917-1918.* Vol. 3. U of Minnesota Press.

Vaswani, A (2017). “Attention is all you need.” In: *Advances in Neural Information Processing Systems.*

Venturini, Miriam (2023). “The imperfect union: labor racketeering, corruption exposure, and its consequences.” In.

Virtanen, Antti et al. (2019). "Multilingual is not enough: BERT for Finnish." In: *arXiv preprint arXiv:1912.07076*.
Vives, Xavier (2005). "Complementarities and games: New developments." In: *Journal of Economic Literature* 43.2, pp. 437–479.
Voigtländer, Nico and Voth, Hans-Joachim (2015). "Nazi indoctrination and anti-Semitic beliefs in Germany." In: *Proceedings of the National Academy of Sciences* 112.26, pp. 7931–7936.
Yanagizawa-Drott, David (2014). "Propaganda and conflict: Evidence from the Rwandan genocide." In: *The Quarterly Journal of Economics* 129.4, pp. 1947–1994.
Zolberg, Aristide R (1972). "Moments of madness." In: *Politics & Society* 2.2, pp. 183–207.

9 Figures and Tables

Figure 1: Excerpt of Wartime Censorship Labels

Ilmoitushinnat:

INKERI

(Pietarin ja Inkerinmaan kuulumisia)

Vastaava toimittaja: Simo Eronen.

Irtonumero 6 kop.

N:o 12 Perjantaina 10 p. (23 p.) Helmikuuta. Konttori: Большая Конюшенная 8, кв. 11. Puh. 509—11 1917

Vilho Sarkasen Asianajotoimisto

Ottaa hoitaakseen asianajotehtäviä Suomessa ja Venäjällä. Asioiden hoito huolellinen ja nopea. Palkkiot kohtuulliset.

Tavattavissa Pietarissa lauantaina klo 1—4 Николаевская № 14 кв. 6.

Telefooni Pietarissa 54—67. S.H.

ZANDER-OPISTO

Ruotsalaista konevoimistelua, hierontaa, värähtelyä, vibratsioni) sähkötystä. Reumatismin, kihdin ja hermotautien erikoislääkintää; sydämen rasvoittumisen, umpitaudin ja unettomuuden parannusta; selkärangan ja köyryselkäisyyden suoristusta; ampuma- y. m. haavojen ja ruhjevammojen hoitoa. Sisustautien vastaanotto arkipäivinä klo 2—6.

Kasanskinkatu 5. Puh. 446-12. S.H.

Osake-Yhtiöitä.

Osakeyhtiö „Granit"

Hanko

Pietari, Fontanka 54. Puh. 175-81.

Hautapatsaita

suuri valikoima varastossa. (S.H.)

Itä-Suomen Telefooni Osakeyhtiö. (S. H.)

Kiviteollisuusyhtiö HÄMÄLÄINEN JA KUMPP.

Pietari, Рѣка Волковка 27. (S.H.)

Koneellinen Kiviveistämö Osakeyhtiö (O. Vainio) (S.H.)

Б. Дворянская 10.

Ruokatavarakauppoja.

N. Muurinen Suomalainen Ruokatavarakauppa (S. H.)

Huveja.

Pietarin suom. voimisteluyhdistys „Riento"

toimeenpanee

Suuren Iltaman

19 p:nä maaliskuuta v. 1917 Kalaschnikovin pörssi-salissa (Poltavskaja 12)

OHJELMA:

I

Torvisoittoa, joht. A. Petroff.

Voimistelua: Vapaaliikkeitä

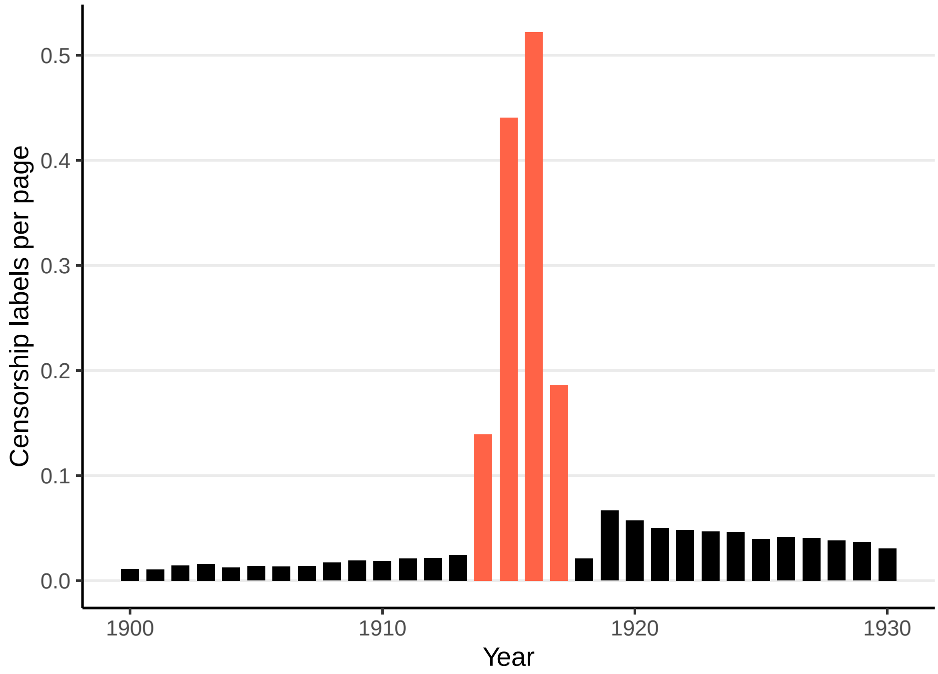
Hyppyjä

Sauvaliikkeitä

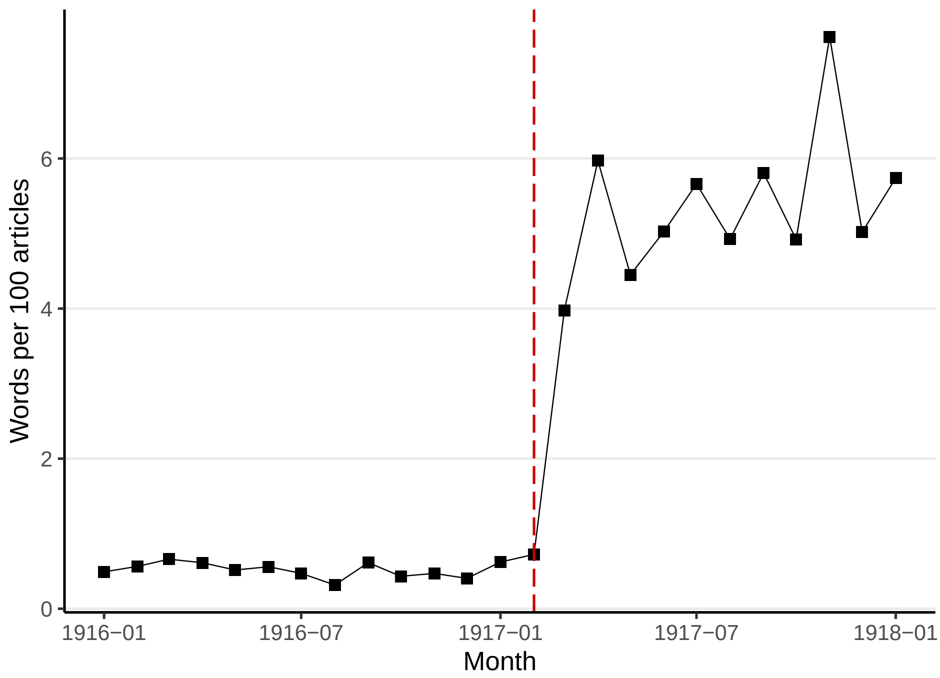
Sowintoon!

Notes: This figure presents an excerpt of wartime censorship labeling in Finnish newspapers. Each news article read and approved by a censor was labelled with the phrase "S.H." (or in Swedish newspapers, "K.C."), which translates "approved by the censor". The issue in question is from newspaper *Inkeri*, published February 10, 1917. The labels are highlighted in red.
Source: digi.nationallibrary.fi

Figure 2: Wartime Censorship and Public Opposition in the Newspapers



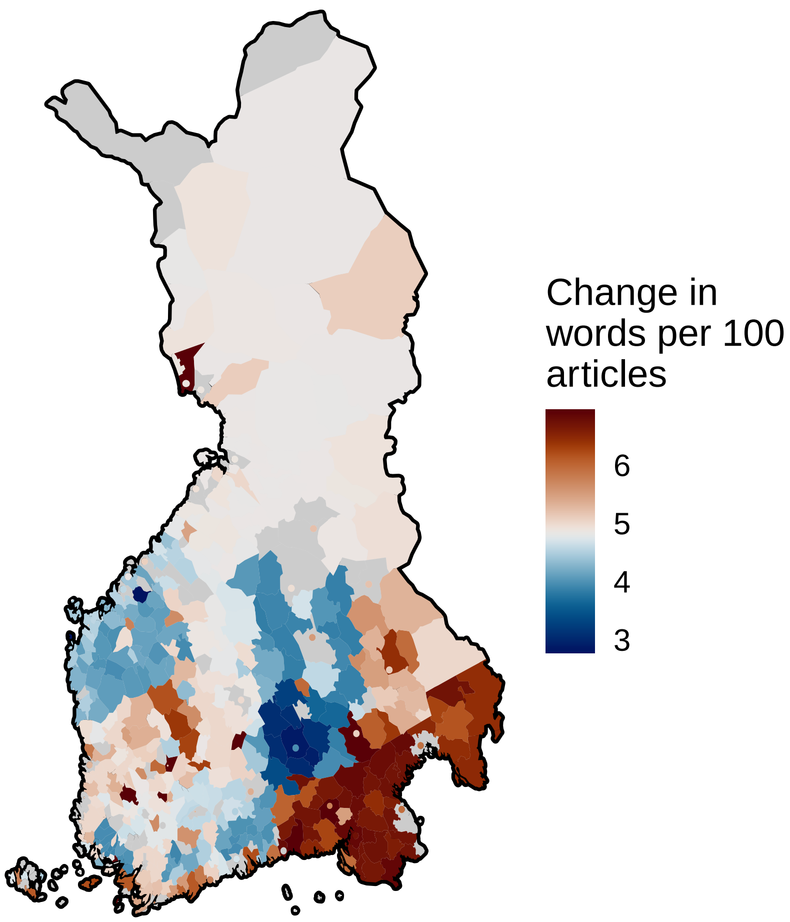
(a) Wartime Censorship



(b) Public Opposition

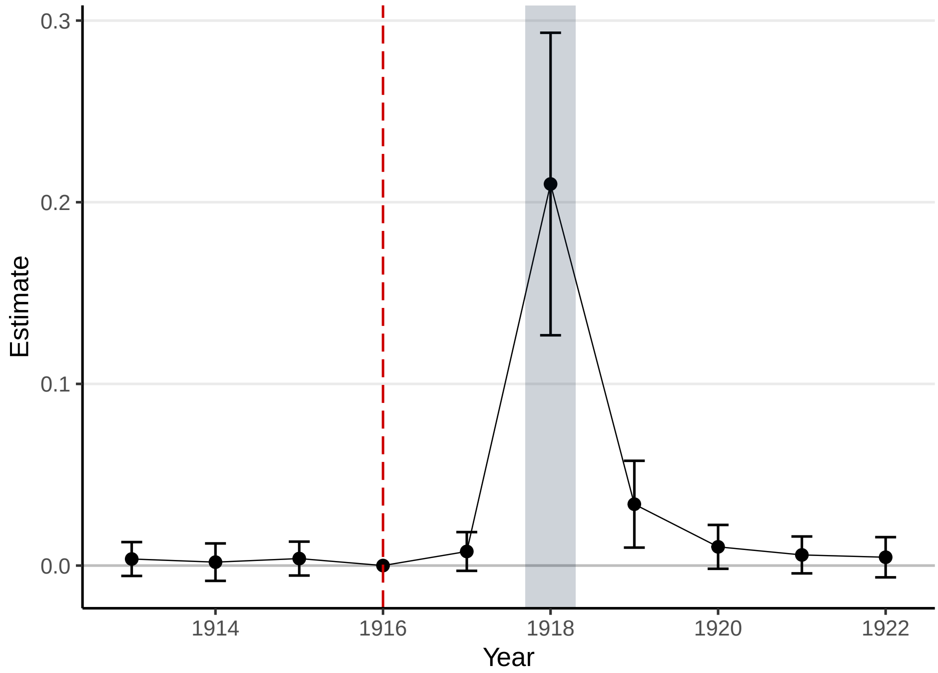
Notes: **Panel (a)**: The figure presents the number of censorship labels per page within the universe of Finnish newspapers. The years of wartime censorship, 1914–1917, are shown in red. **Panel (b)**: The figure presents the average term frequency of inflammatory words within the universe of Finnish newspapers in each month. The set of inflammatory words are: "revolution", "democracy", "anarchy", "oppression", "freedom" and "*butcher*". The red vertical line marks the month preceding the abolition of censorship.
Source: `digi.nationallibrary.fi`

Figure 3: Spatial Distribution of Local Change in Public Opposition

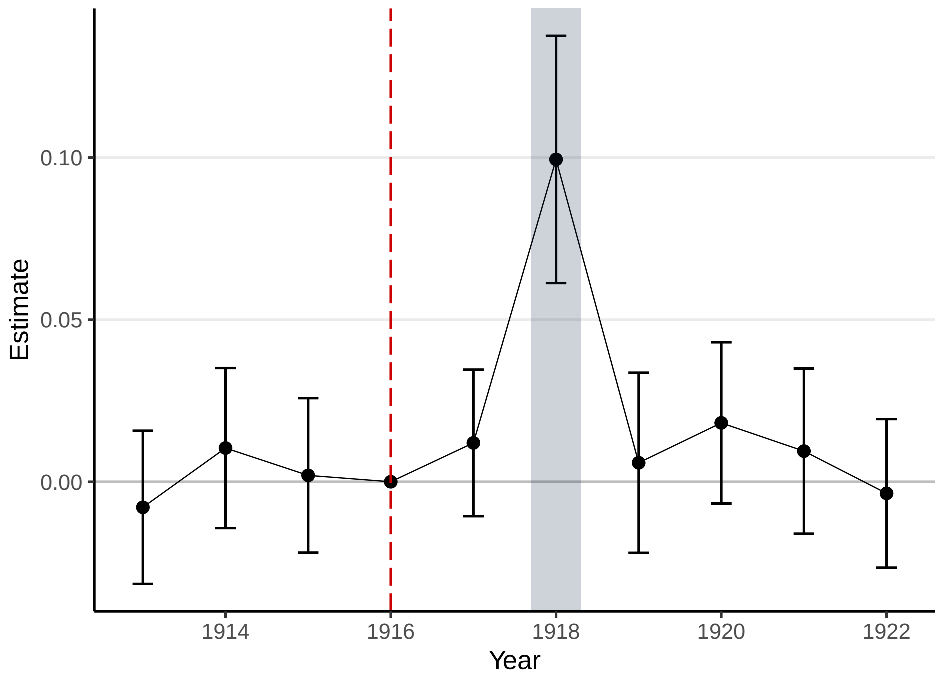


Notes: This graph presents the local change in public opposition across municipalities after the abolition of censorship, as defined in equation (1).

Figure 4: Event-Study Estimates of a Change in Public Opposition on War Mortality and Overall Mortality



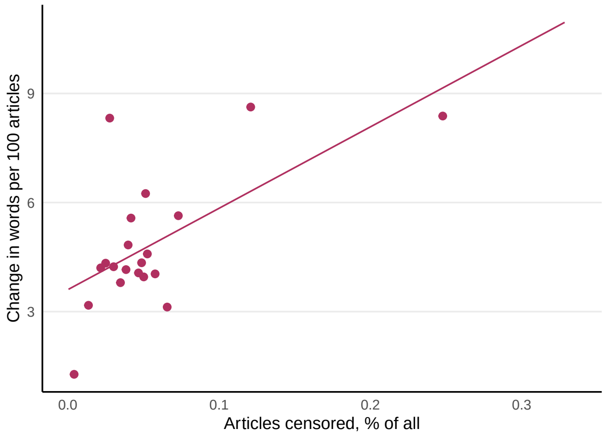
(a) War Mortality



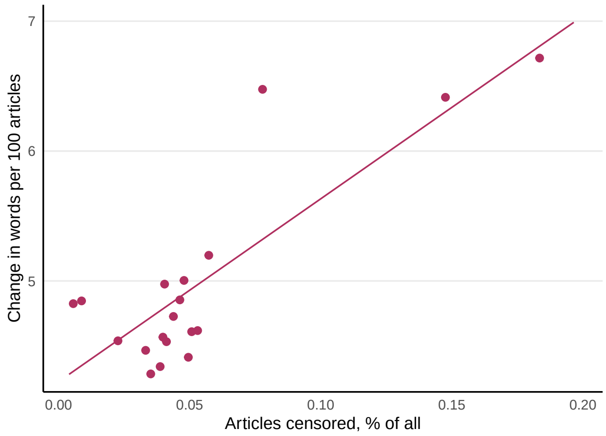
(b) Overall Mortality

Notes: The figure presents event-study estimates of β_τ from equation (3). In Panel (a), the outcome is the log of war mortality, and in Panel (b) it is the log of overall mortality. The model also includes municipality and year fixed effects, as well as baseline controls for SDP vote share, historical mobilization, log population in 1916, longitude and latitude, each interacted with year effects. The red vertical line indicates the year preceding the abolition of censorship. Standard errors are clustered at municipal level. The error bars present 95% confidence intervals.

Figure 5: Relationship Between Censorship and the Change in Public Opposition



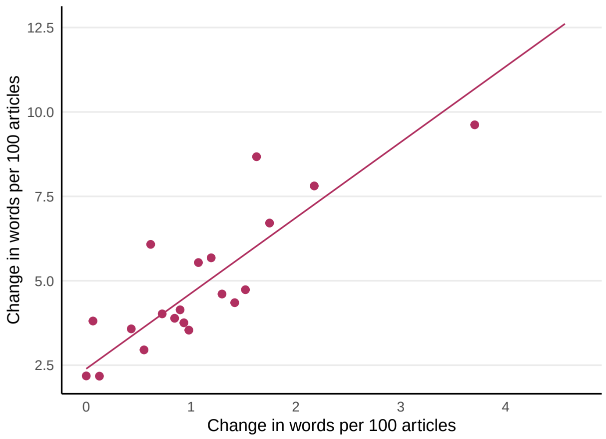
(a) Newspaper-level



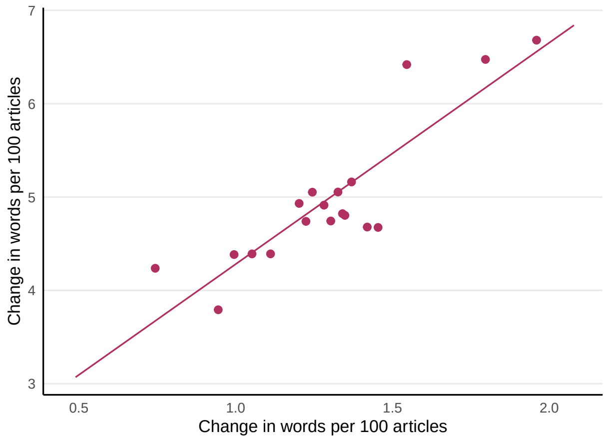
(b) Municipal level

Notes: **Panel (a)**: The figure shows binscatter of the fraction of censored articles and the change in public opposition. The unit of observation is journal. **Panel (b)**: The figure shows binscatter of the fraction of censored articles and the change in public opposition. The unit of observation is municipality. The model also includes baseline controls for SDP vote share, historical mobilization, log population in 1916, longitude and latitude.

Figure 6: Relationship Between News Coverage of Russian Revolution and the Change in Public Opposition



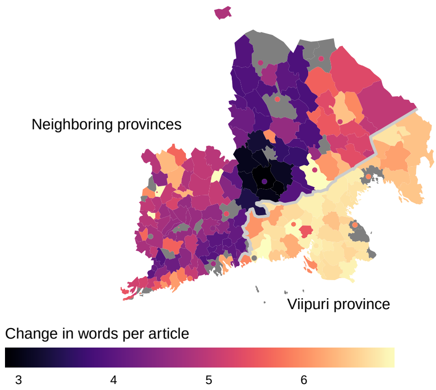
(a) Newspaper-level



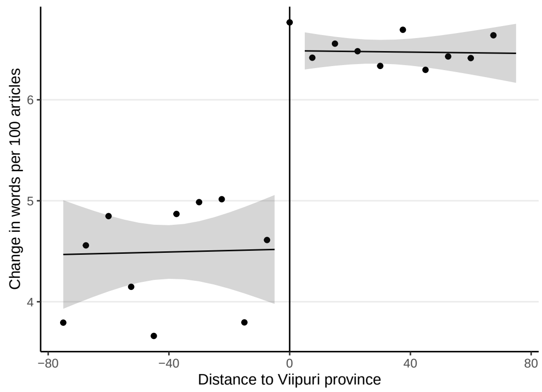
(b) Municipal level

Notes: **Panel (a)**: The figure shows binscatter of the news coverage of Russian Revolution and the change in public opposition. The unit of observation is journal. **Panel (b)**: The figure shows binscatter of the news coverage of Russian Revolution and the change in public opposition. The unit of observation is municipality. The model also includes baseline controls for SDP vote share, historical mobilization, log population in 1916, longitude and latitude.

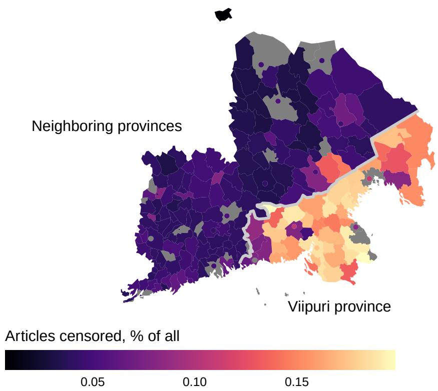
Figure 7: First Stage RD Plots



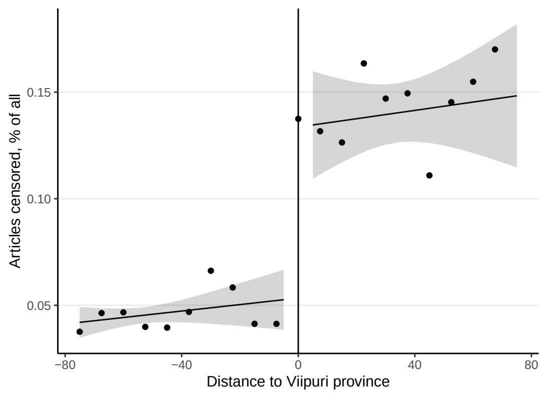
(a) Change in Public Opposition in 3D



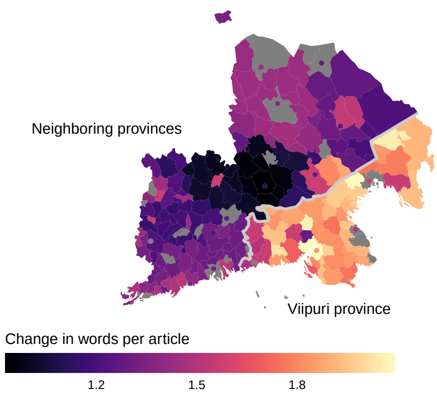
(b) Change in Public Opposition in 2D



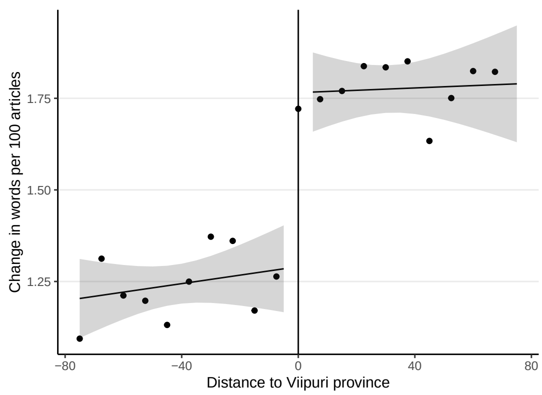
(c) Censorship in 3D



(d) Censorship in 2D



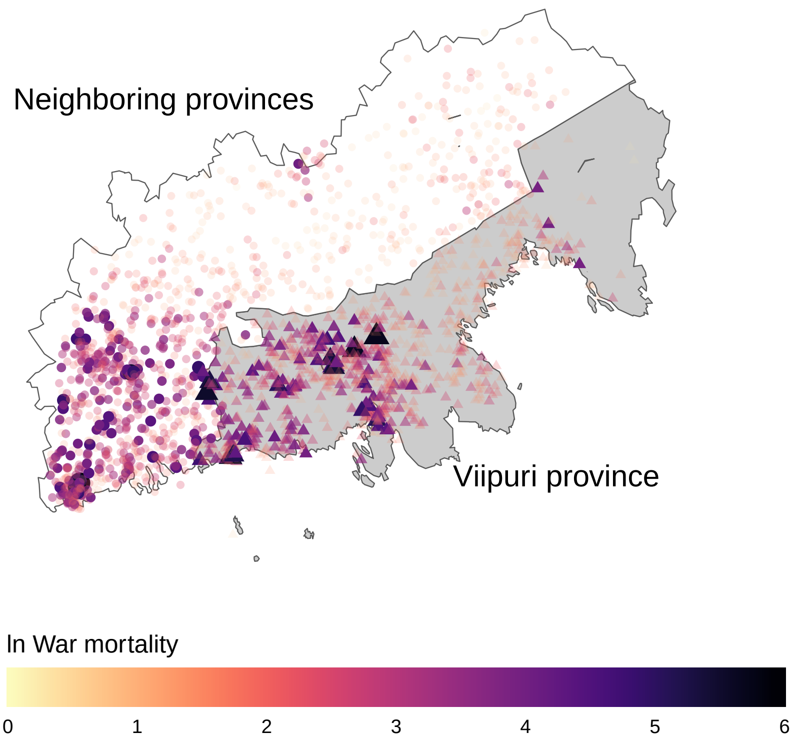
(e) News Coverage of Russian Revolution in 3D



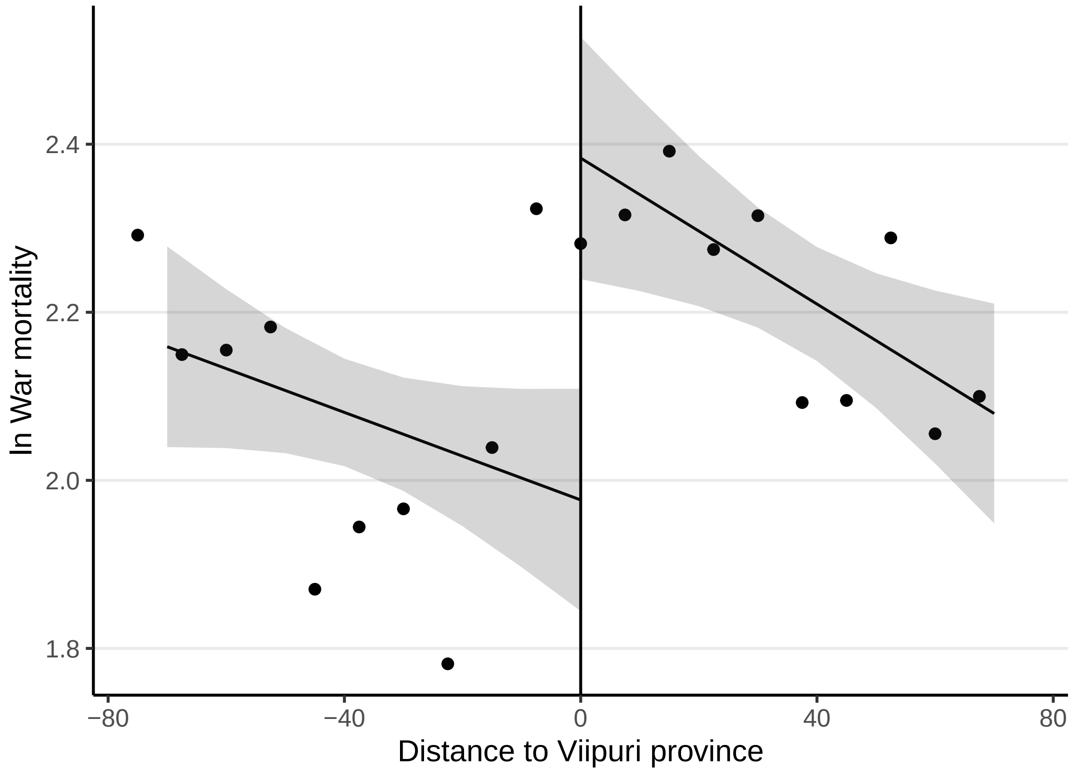
(f) News Coverage of Russian Revolution in 2D

Notes: The sub-figures in each row present first stage RD plots in three and two dimensions for each independent variable. The unit of observation is municipality. Each sub-figure on the right presents the mean value of each main treatment variable in twenty equally sized bins along the distance to Viipuri province border. Standard errors are clustered at municipal level. The gray ribbons present 95% confidence intervals.

Figure 8: Reduced Form RD Plots



(a) War Mortality in 3D



(b) War Mortality in 2D

Notes: **Panel (a)**: The figure shows reduced form RD plot in three dimensions. The unit of observation is village. **Panel (b)**: The figure shows reduced form RD plot in two dimensions. The unit of observation is village, and standard errors are clustered at the village level. The gray ribbons present 95% confidence intervals.

Table 1: Covariate Balance by Change in Public Opposition

	Mean	Observations	Coefficient
SDP vote share	0.44	415	−0.03**
			(0.01)
Mobilization 1905	0.19	437	0.02
			(0.02)
ln Workers houses	1.93	437	0.01
			(0.04)
ln Strikers	57.02	437	0.14
			(0.10)
Land gini	0.60	395	−0.01
			(0.01)
Shortage	1.04	437	−0.02
			(0.04)
Swedish population, % of all	0.12	436	0.02
			(0.01)
ln Population 1916	6827.57	437	0.05
			(0.04)
ln Slope	4.04	436	−0.04***
			(0.01)

Notes: The unit of observation is a municipality. All regressions control for longitude, latitude and log population, except the regression with log population as the outcome, which controls for longitude and latitude. Standard errors are clustered at municipal level. *** $p < 0.01$,** $p < 0.05$,* $p < 0.1$.

Table 2: Change in Public Opposition and Mortality in 1918

	ln War mortality		ln Mortality	
	(1)	(2)	(3)	(4)
Change in public opposition × Post	0.21***		0.10***	
	(0.04)		(0.02)	
Above median change in public opposition × Post		0.21***		0.09***
		(0.07)		(0.03)
Municipality fixed effects	✓	✓	✓	✓
Time fixed effects	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Observations	2,490	2,490	2,486	2,486
Mean Y	2.81	2.81	18.55	18.55
R^2	0.92	0.93	0.67	0.68

Notes: Difference-in-differences estimates from equation (2). The unit of observation is a municipality. War mortality is the number of war casualties per 1000 people +1. Mortality is the number of overall deaths per 1000 people +1. All regressions control for municipal fixed effects, time fixed effects, as well as baseline controls for SDP vote share, historical mobilization, log population and longitude and latitude. Standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Change in Public Opposition and Various Mobilization Outcomes in 1918

	OLS						Cox PH
	ln Red mortality	ln White mortality	Red casualties, % of all	Red Guard casualties, % of Reds	White Guard casualties, % of Whites	ln Investigations per capita	$h(t)$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Change in public opposition	0.26***	0.02	0.05***	0.05***	0.005	0.11***	0.06***
	(0.05)	(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.005)
Controls	✓	✓	✓	✓	✓	✓	✓
Observations	415	415	414	364	388	415	415
Mean Y	9.45	2.62	0.55	0.44	0.30	1.59	
R^2	0.62	0.11	0.55	0.28	0.05	0.26	

Notes: The unit of observation is a municipality. Red mortality is the number of Red casualties per 1000 people +1. White mortality is the number of White casualties per 1000 people +1. Investigations per capita is the sum of post-conflict investigations of the rebels per 1000 people +1. $h(t)$ is the hazard function. All regressions include baseline controls for SDP vote share, historical mobilization, log population and longitude and latitude. Standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4: Censorship and the News Coverage of Russian Revolution

	Change in public opposition				ln War mortality	
	(1)	(2)	(3)	(4)	(5)	(6)
Censorship	8.0***	5.2***				
	(1.4)	(0.38)				
News on Russian Revolution			2.3***	2.5***		
			(0.32)	(0.15)		
Censorship × Post					0.77**	
					(0.37)	
News on Russian Revolution × Post						0.88***
						(0.14)
Municipality fixed effects					✓	✓
Time fixed effects					✓	✓
Controls	✓	✓	✓	✓	✓	✓
Observations	121	415	121	415	2,490	2,490
R^2	0.20	0.46	0.50	0.55	0.92	0.93

Notes: The unit of observation in columns (1) and (3) is a journal, in columns (2), (4), (5) and (6) it is a municipality. Censorship is the average number of censorship labels per article, normalized by dividing by its maximum value, so that it varies between zero and one. News on Russian Revolution is the average term frequency of the words "bolsheviks", "mensheviks", "Lenin" and "Trotski" after the abolition of censorship. Standard errors are clustered at journal level in columns (1) and (3), and at municipal level in columns (2), (4), (5) and (6). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

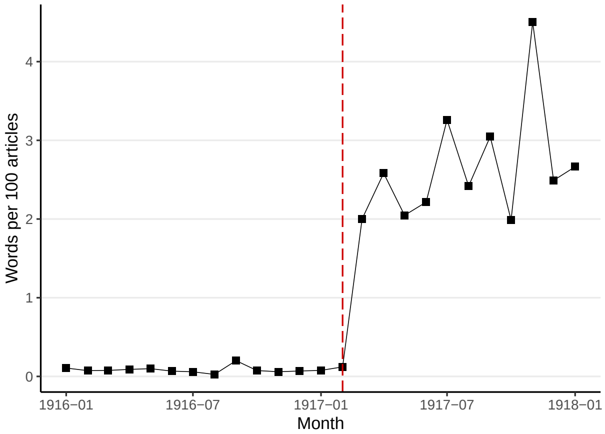
Table 5: Change in Public Opposition and Mortality in 1918: Results from RDD

	ln War mortality					
	Lat-Lon	Dist	Lat-Lon & Dist	Quadratic Lat-Lon	Baseline controls	Segment FE
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Public opposition						
Change in public opposition	0.18***	0.24***	0.14***	0.24***	0.27***	0.22***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Observations	1408	1408	1408	1408	1379	1379
First-stage F	974.33	846.64	1047.7	1033.6	751.98	1145.8
Panel B: Reduced form						
Viipuri province	0.38***	0.41***	0.32***	0.48***	0.50***	0.48***
	(0.07)	(0.10)	(0.09)	(0.08)	(0.08)	(0.09)
Observations	1540	1540	1540	1540	1505	1505

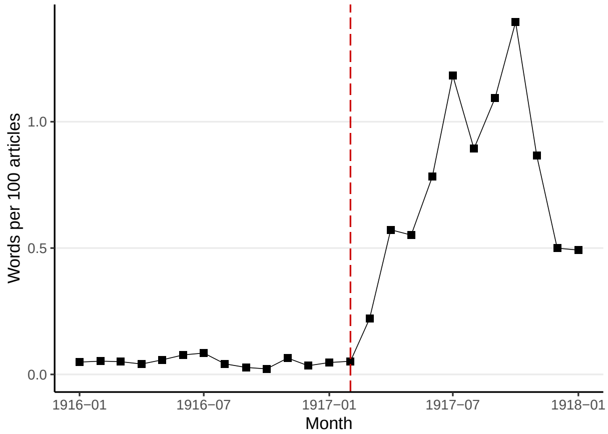
Notes: Fuzzy RD estimates from equation (6). The unit of observation is a village. Columns (1) and (3) through (6) include a linear polynomial in coordinates. Column (2) and (3) include a linear polynomial in distance to the Viipuri province border. Column (4) includes a quadratic polynomial in coordinates. Column (5) and (6) include the baseline controls from main analysis: SDP vote share, historical mobilization and log population. Column (6) adds segment fixed effects. Standard errors are clustered at village level. *** $p < 0.01$,** $p < 0.05$,* $p < 0.1$.

A Additional Figures and Tables

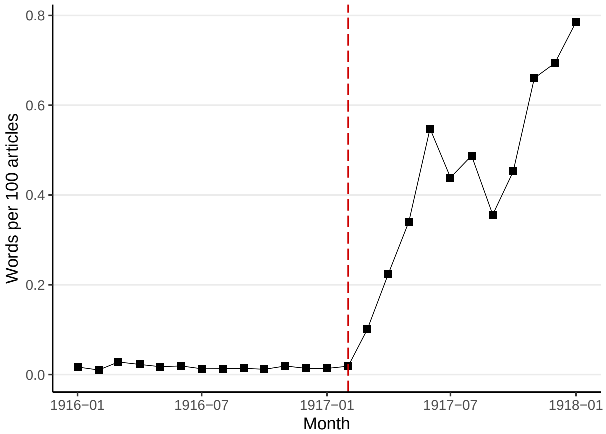
Figure A1: Public Opposition in the Newspapers by Word



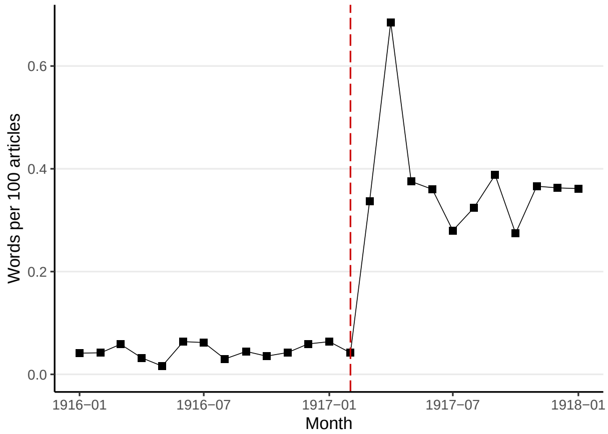
(a) Word: Revolution



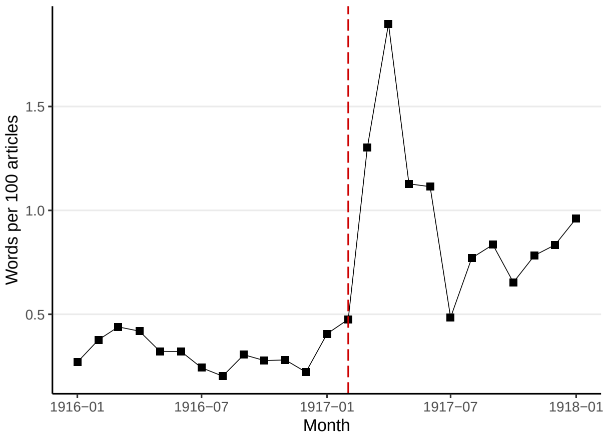
(b) Word: Democracy



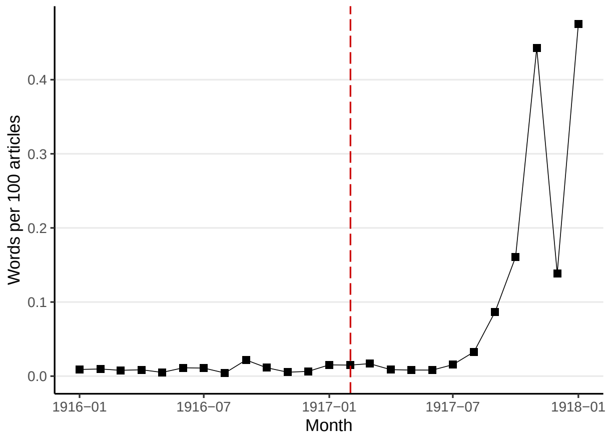
(c) Word: Anarchy



(d) Word: Oppression



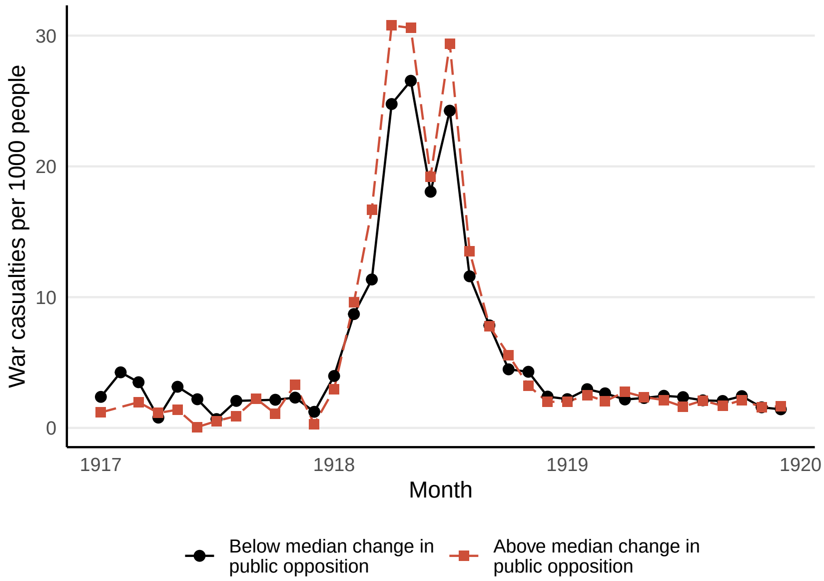
(e) Word: Freedom



(f) Word: *Butcher*

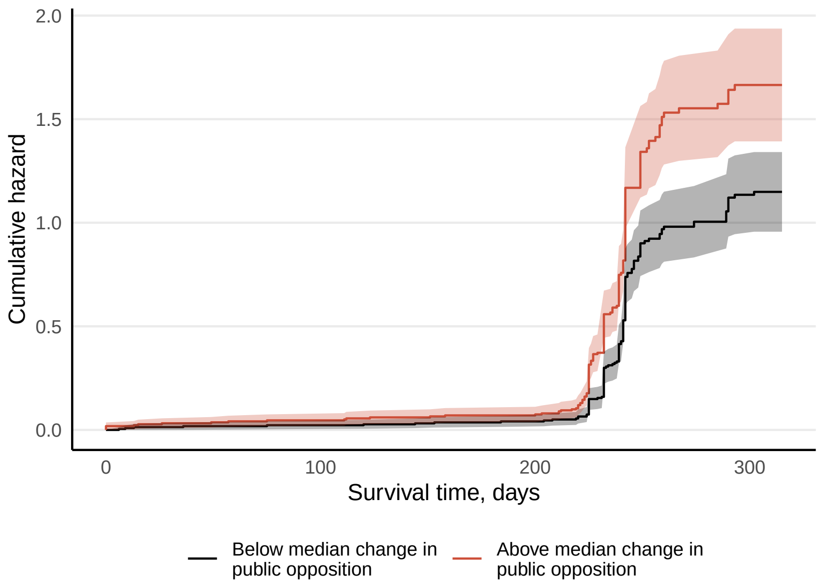
Notes: Each sub-figure presents the average term frequency of the given word in each month. The red vertical line marks the month preceding the abolition of censorship.

Figure A2: War Mortality for Municipalities With Above Median and Below Median Change in Public Opposition



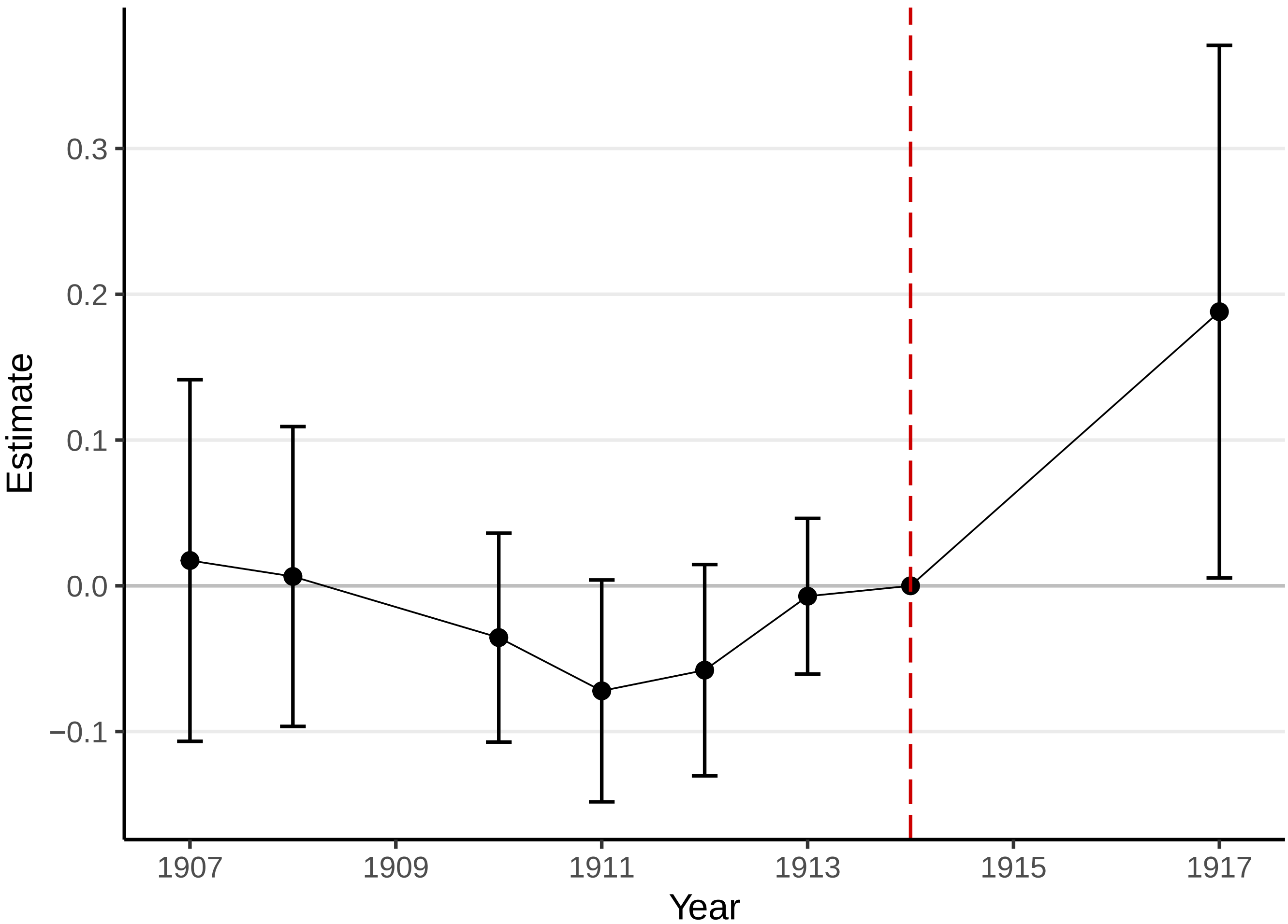
Notes: This figure presents raw monthly war mortality data for municipalities with above median and below median change in public opposition. The numbers have been multiplied by twelve to reflect year-equivalent levels.

Figure A3: Foundation of Red Guards: Cumulative Hazard Estimates for Municipalities With Above Median and Below Median Change in Public Opposition



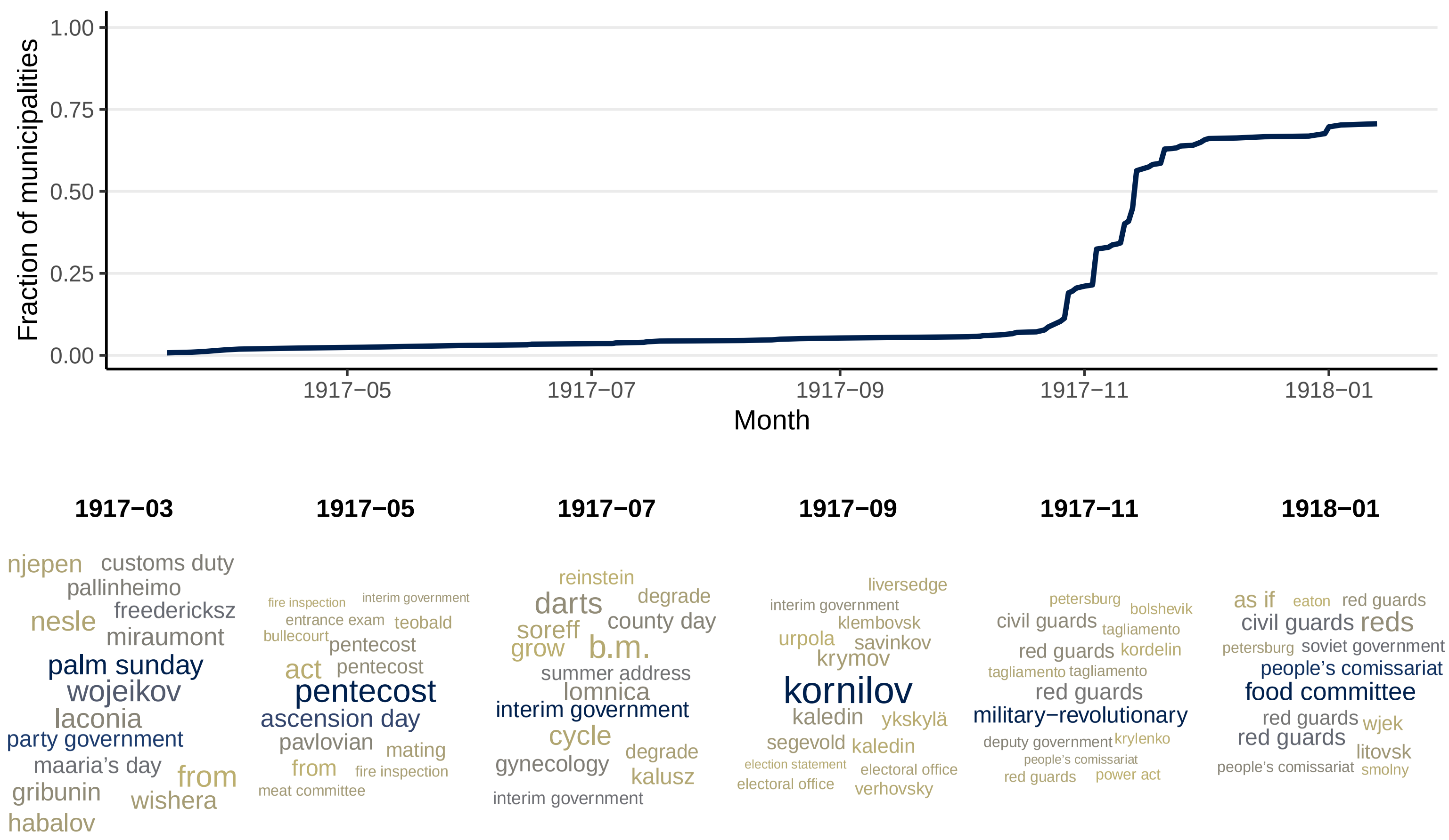
Notes: The figure presents Nelson-Aalen estimates of cumulative hazard functions $\hat{H}(t)$ from equation (7). The estimates are computed separately for municipalities with above median and below median change in public opposition.

Figure A4: Event-Study Estimates of a Change in Public Opposition on Striking



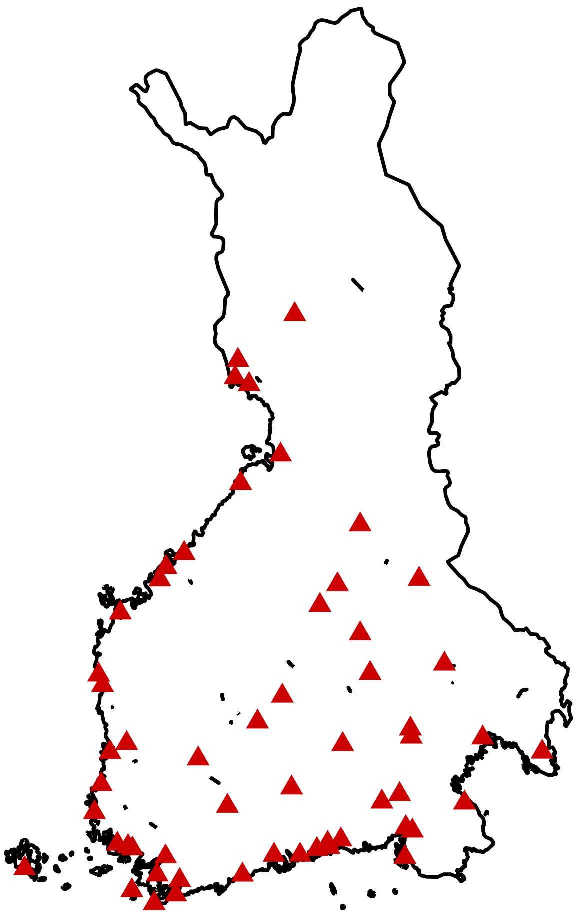
Notes: The figure presents event-study estimates of β_τ from equation (3). The outcome is log of strikers per capita +1. The model also includes municipality and year fixed effects, as well as baseline controls for SDP vote share, historical mobilization, log population in 1916, longitude and latitude, each interacted with year effects. The red vertical line indicates the last observation preceding the abolition of censorship. Standard errors are clustered at municipal level. The error bars present 95% confidence intervals.

Figure A5: Foundation of Red Guards and Trending Terms



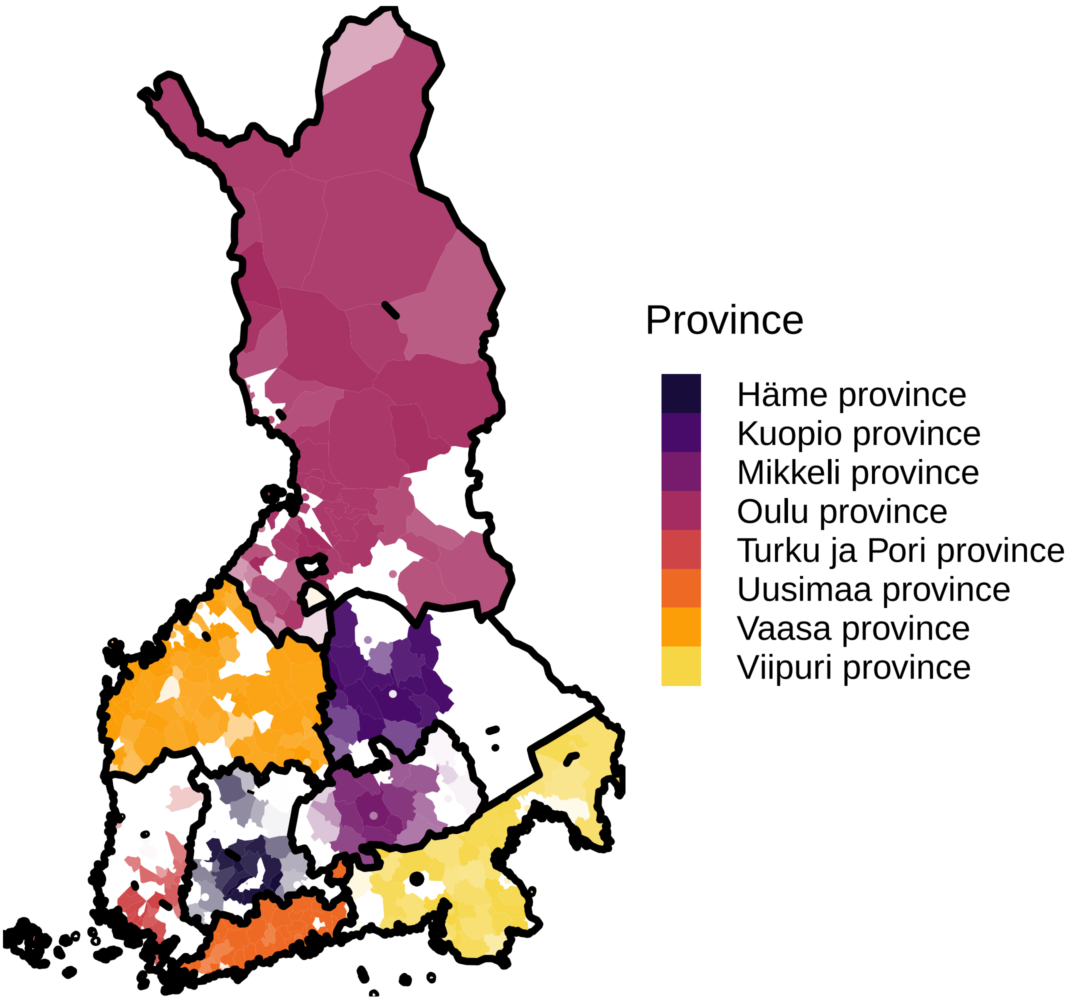
Notes: This figure shows the fraction of municipalities with a Red Guard alongside the most trending news terms for selected months. The news terms are identified using TF-IDF. The larger or darker the font, the more trending the word.

Figure A6: Spatial Distribution of Russian Telegraph Stations in Finland



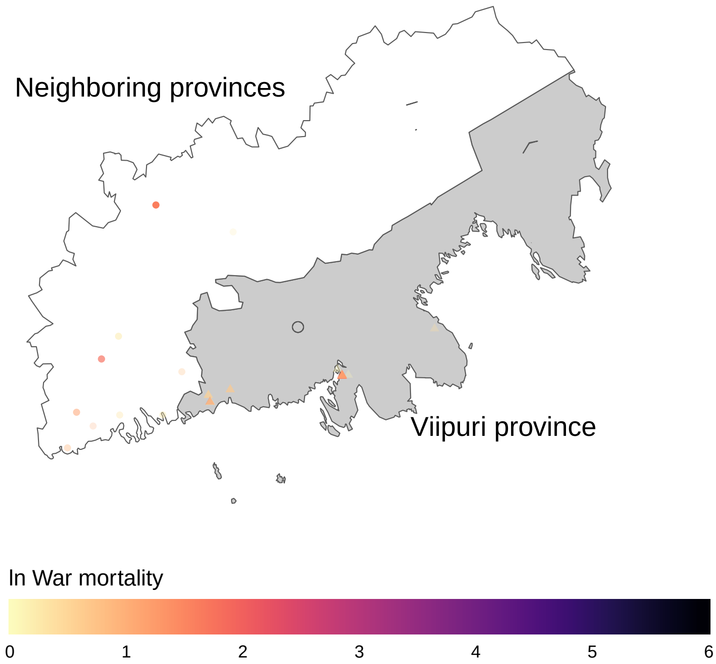
Notes: This figure shows the locations of Russian telegraph stations in Finland in 1899.
Source: *Suomen Kartasto 1899.*

Figure A7: Geographic Patterns of Local News Consumption

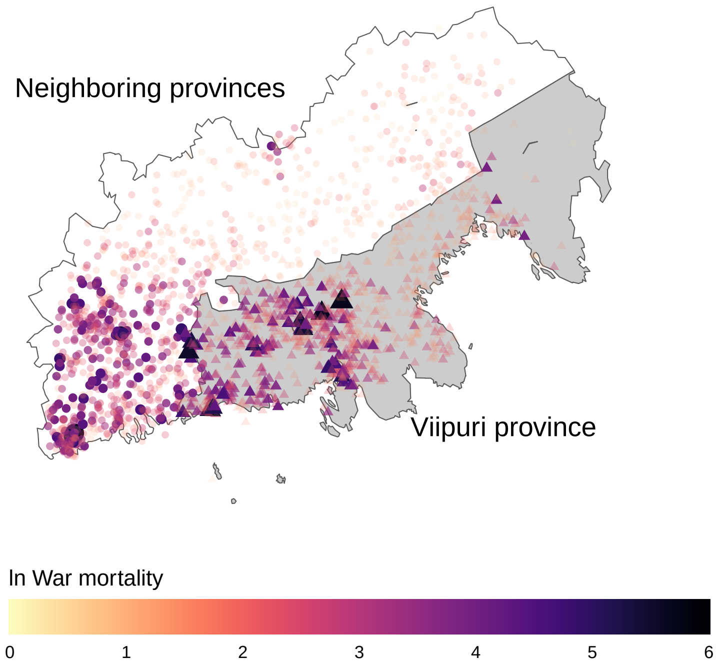


Notes: This figure shows the local saturation of news markets across municipalities. Local papers are defined as papers published within province boundaries, superimposed in black lines. Each province has different local papers, and thus different color gradient. The darker the color, the greater the fraction of news consumption from local papers.

Figure A8: Reduced Form RD Plots Before and After the Abolition of Censorship



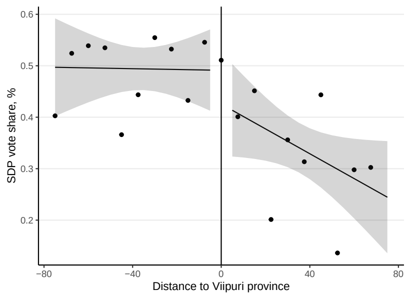
(a) War Mortality Before March 20, 1917



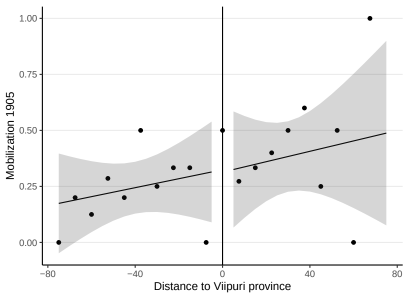
(b) War Mortality After March 20, 1917

Notes: **Panel (a)**: The figure shows reduced form RD plot in three dimensions before the abolition of censorship in March 20, 1917. The unit of observation is village. **Panel (b)**: The figure shows reduced form RD plot in three dimensions after the abolition of censorship in March 20, 1917. The unit of observation is a village.

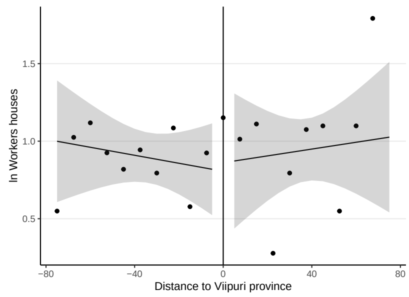
Figure A9: Covariate Balance across Viipuri Province Border



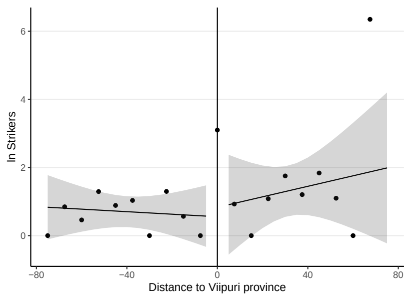
(a) SDP Vote Share



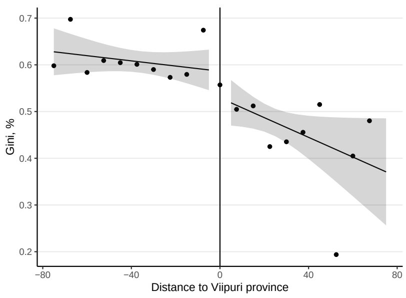
(b) Mobilization 1905



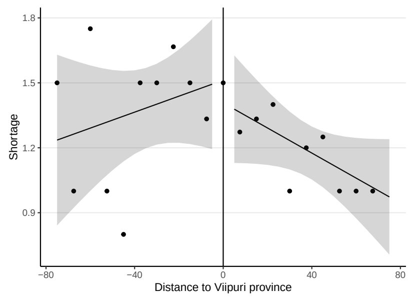
(c) ln Workers Houses



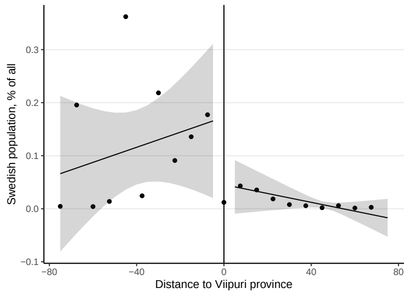
(d) ln Strikers



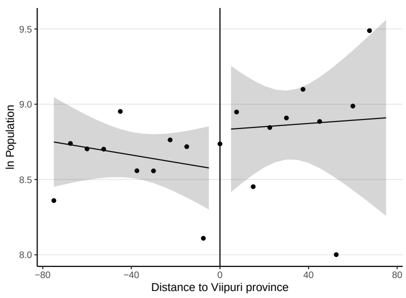
(e) Land Gini



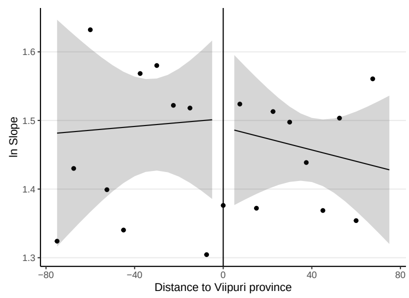
(f) Shortage



(g) Swedish Population, % of All



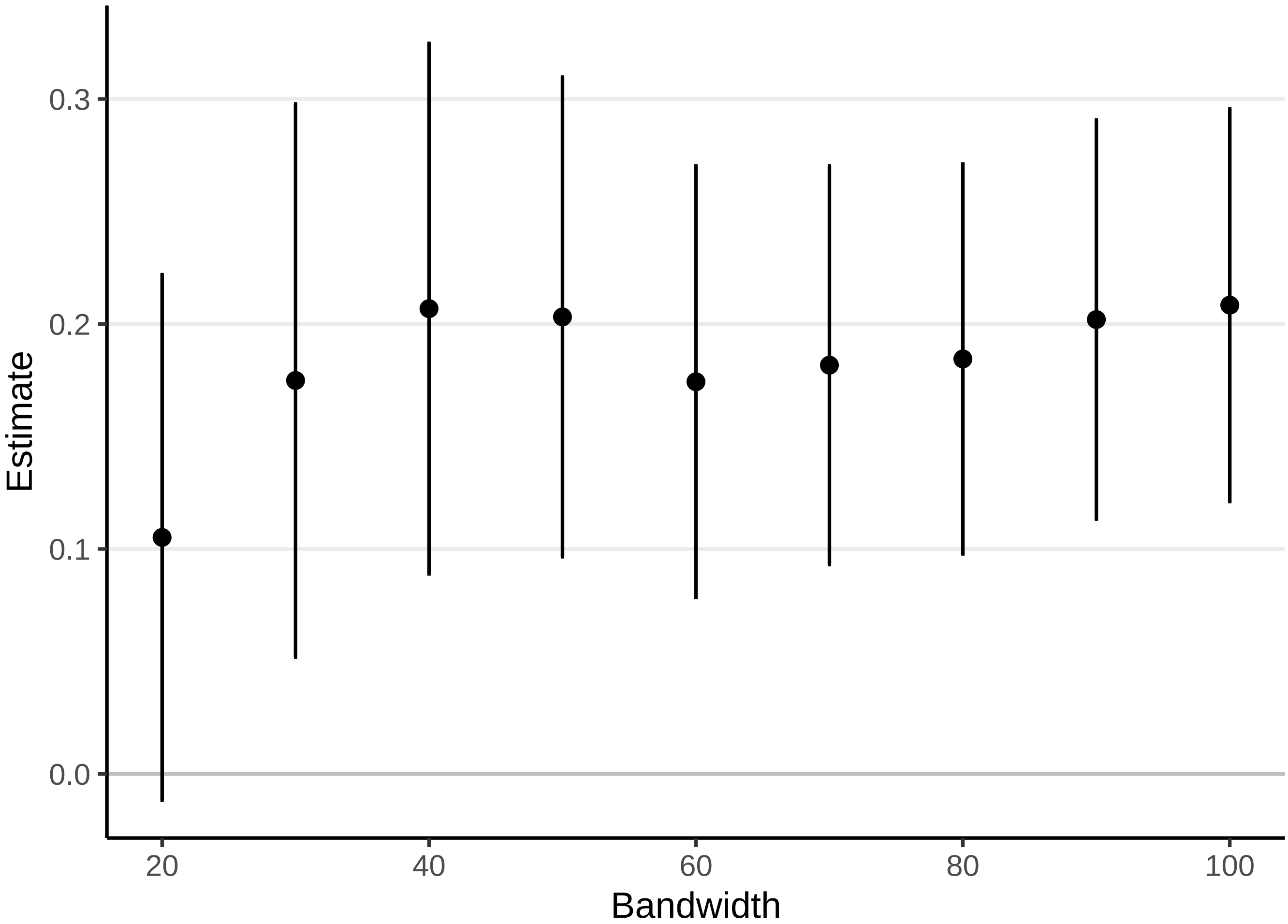
(h) ln Population 1916



(i) ln Slope

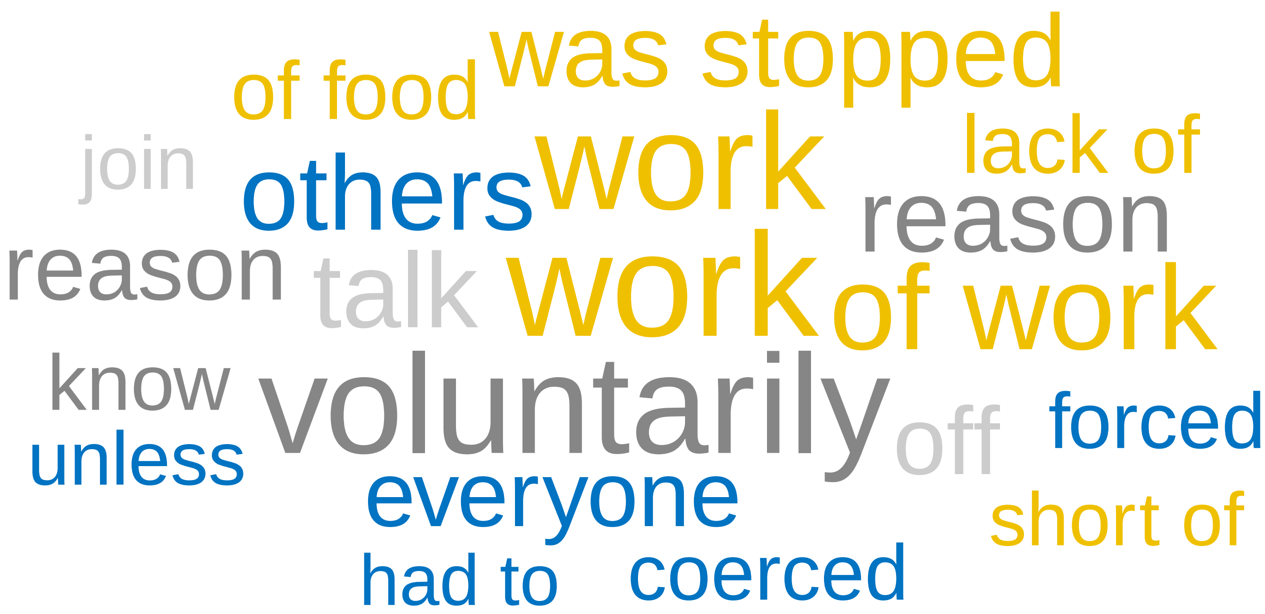
Notes: Each sub-figure presents the mean value of each covariate in twenty equally sized bins along the distance to Viipuri province border. Standard errors are clustered at municipal level. The gray ribbons present 95% confidence intervals.

Figure A10: Bandwidth Sensitivity of the Baseline RD Specification



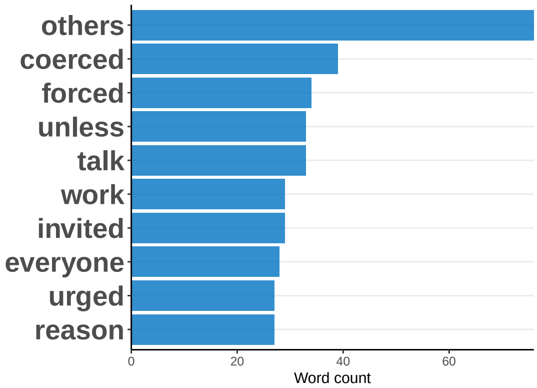
Notes: This graph shows sensitivity of the baseline RD specification (6) to different bandwidth lengths. The RD polynomial is local linear polynomial in coordinates. The model includes no additional controls. The error bars present 95% confidence intervals.

Figure A11: Word Cloud of the Most Common Words Describing Mobilization Motives

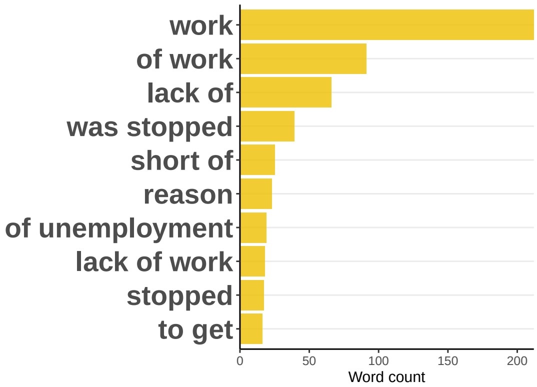


Notes: This graph illustrates the most common words describing mobilization motives in the interrogation records. The larger the font, the more common the word. Yellow words reflect material, dark grey words ideological and blue words conformity motives. Words in light grey could not be connected to any of the above categories.

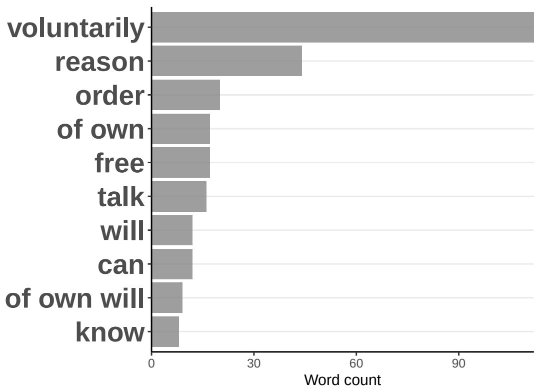
Figure A12: Most Common Words in Each Motive Category



(a) Conformity



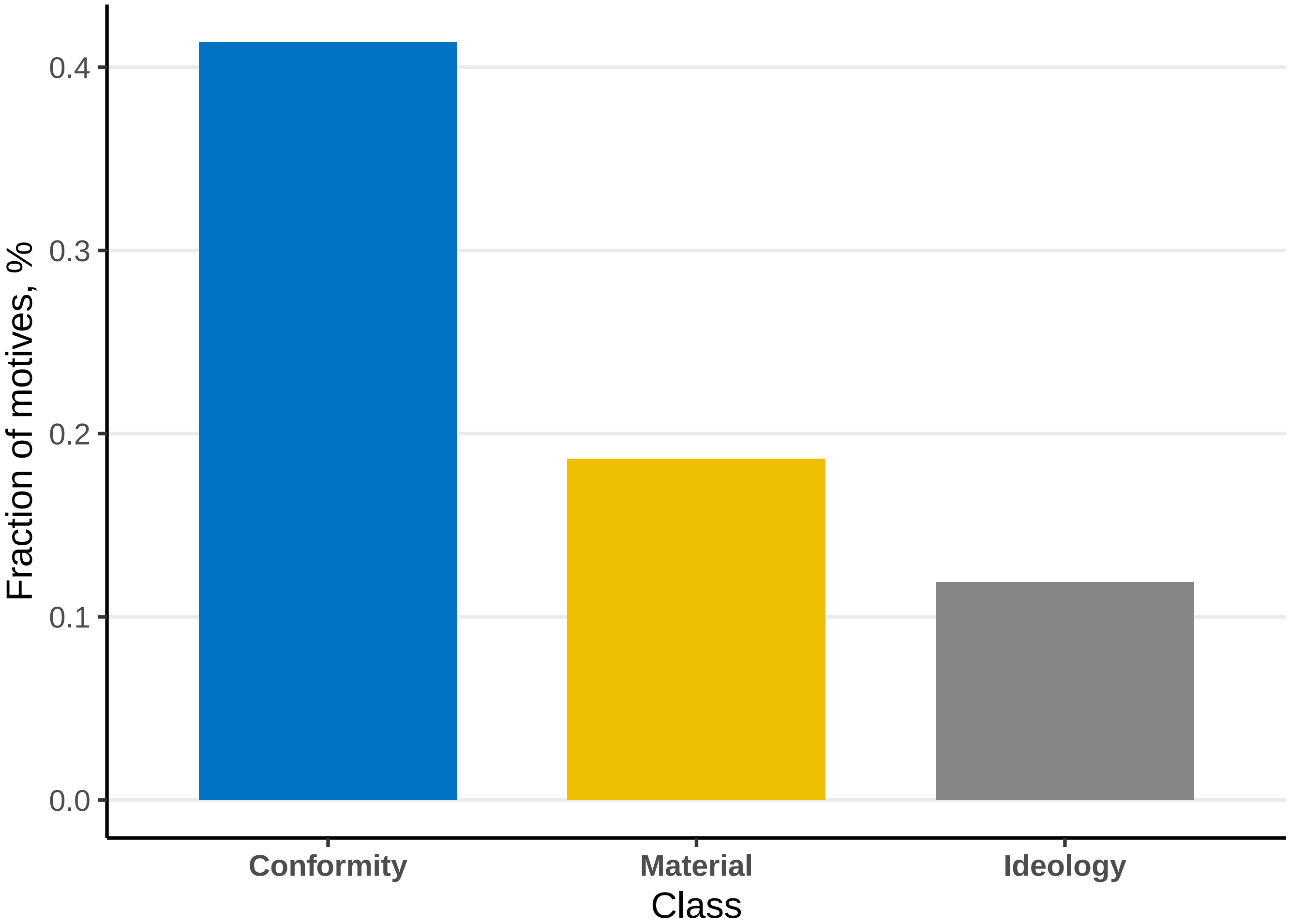
(b) Material



(c) Ideology

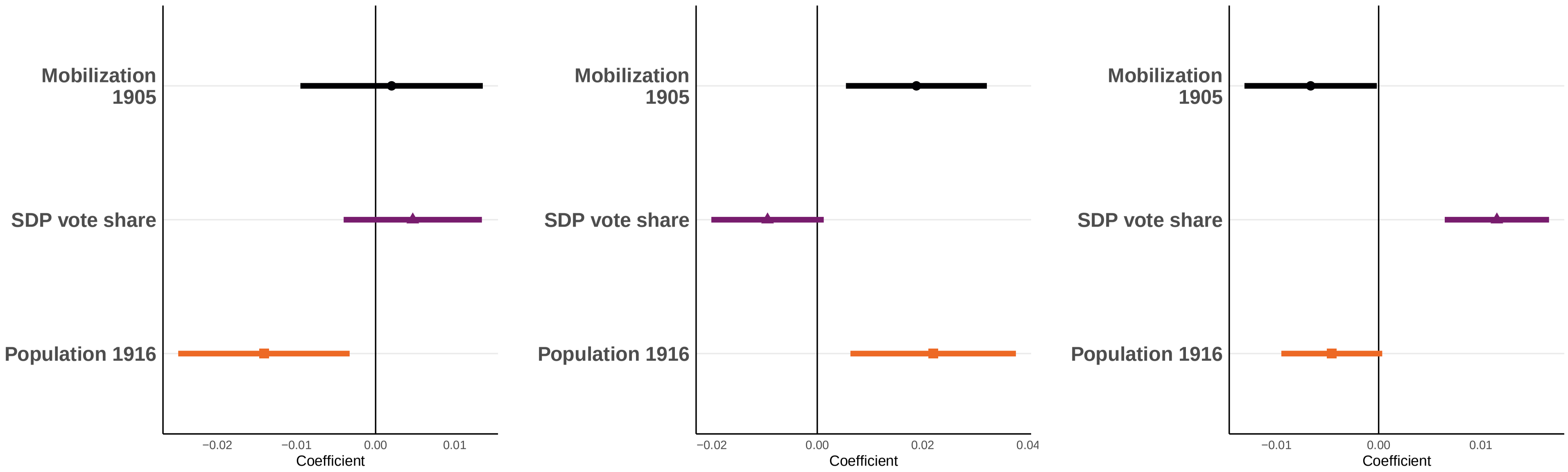
Notes: The sub-figures in each row present most frequent words for each motive category in the training data.

Figure A13: Share of Motives Containing Elements of Conformity, Ideology or Material Reasons



Notes: This graph shows the share of motives that contain elements of conformity, ideology or material justification, respectively. The motives are classified by estimating a multiclass text classification model using a feed-forward neural net.

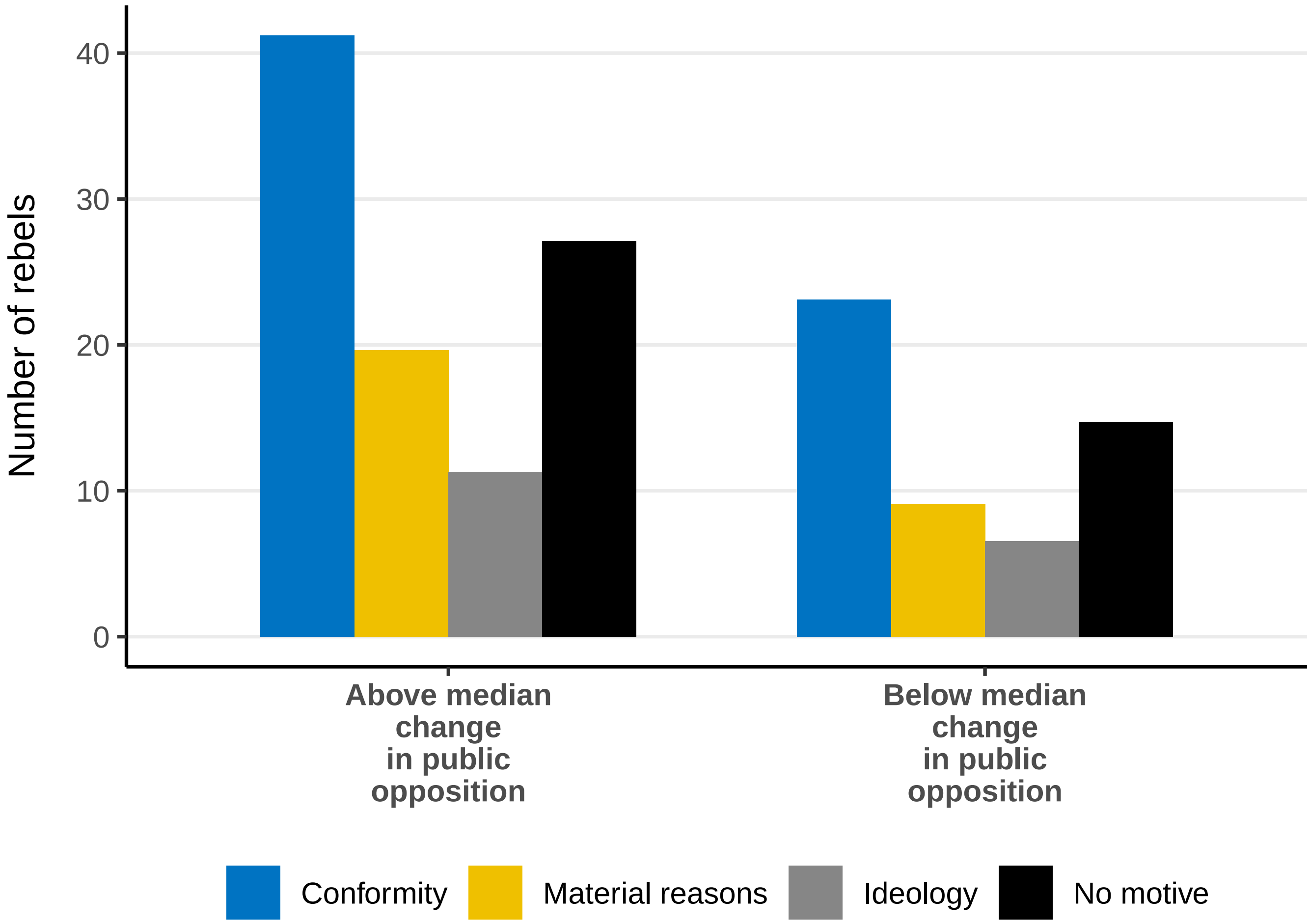
Figure A14: Relationship Between Individual Mobilization Motives and Key Covariates



(a) Conformity **(b)** Material **(c)** Ideology

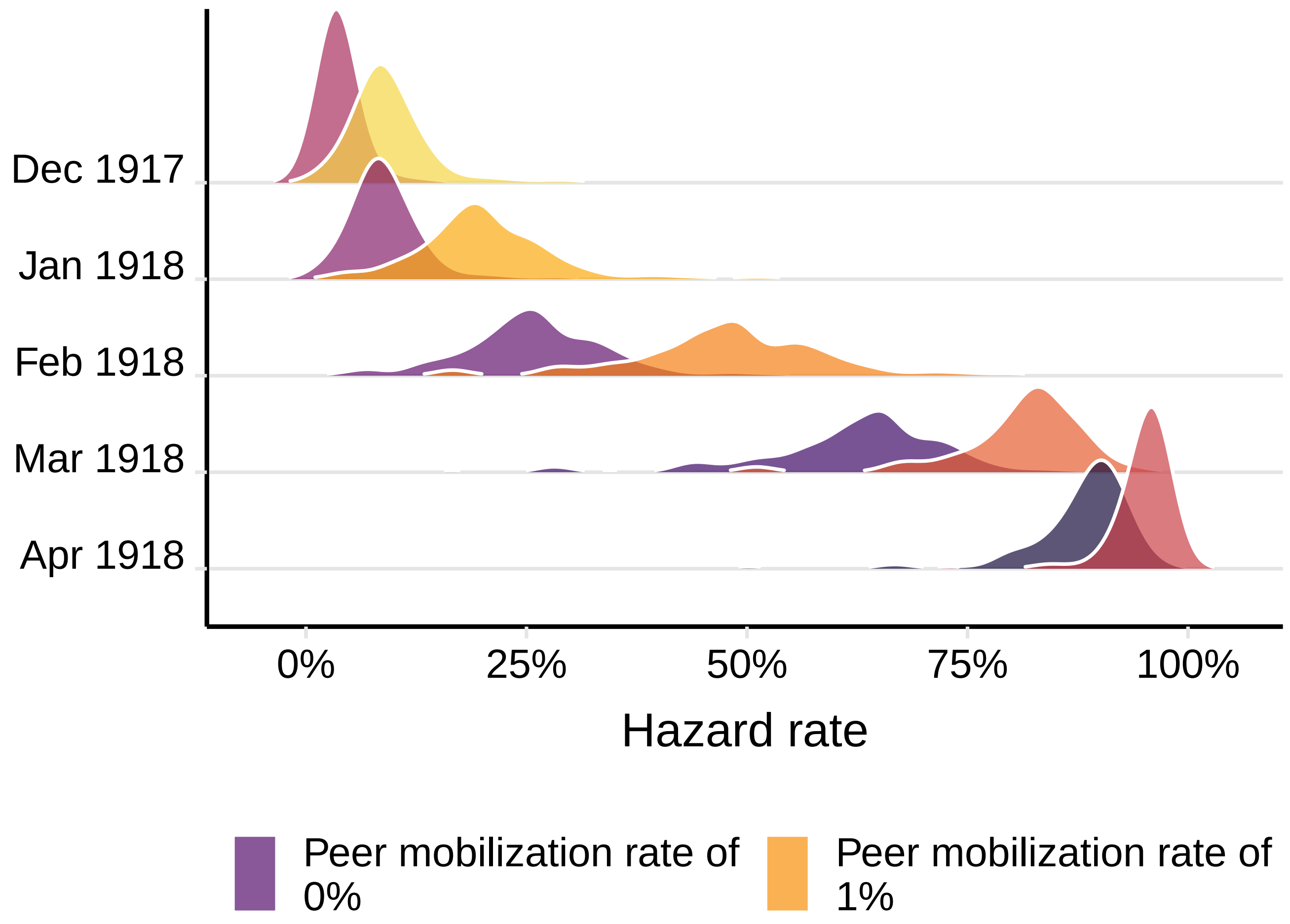
Notes: This graph shows cross-sectional estimates from equation (8). In Panel (a), the outcome is an indicator variable, which gets value one if the individual appealed to conforming motives for joining the rebellion, and is zero otherwise. In Panels (b) and (c), the outcome is defined similarly for ideological and material motives. The model also includes baseline controls for longitude and latitude. Standard errors are clustered at municipal level. The error bars present 95% confidence intervals.

Figure A15: Number of Differently Motivated Rebels for Municipalities With Above Median and Below Median Change in Public Opoosition



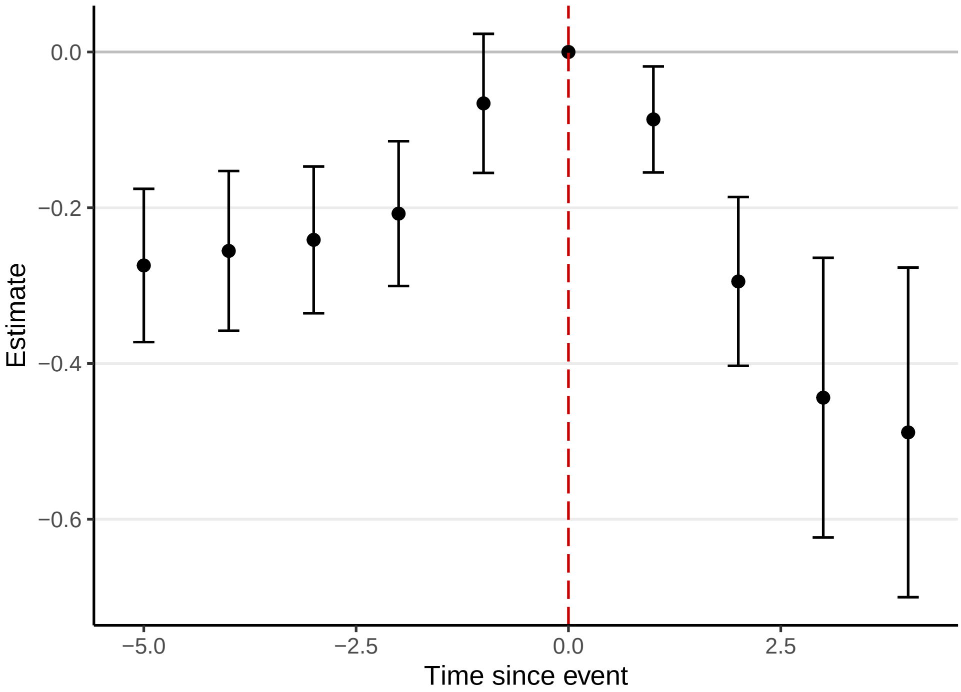
Notes: This graph shows the number of rebel motives that contain elements of conformity, ideology, material justification, or none of the above for municipalities with above and below median change in public opposition. The motives are classified by estimating a multiclass text classification model using a feed-forward neural net.

Figure A16: Predicted Hazards for Rebels in Municipalities With Different Peer Mobilization Rates

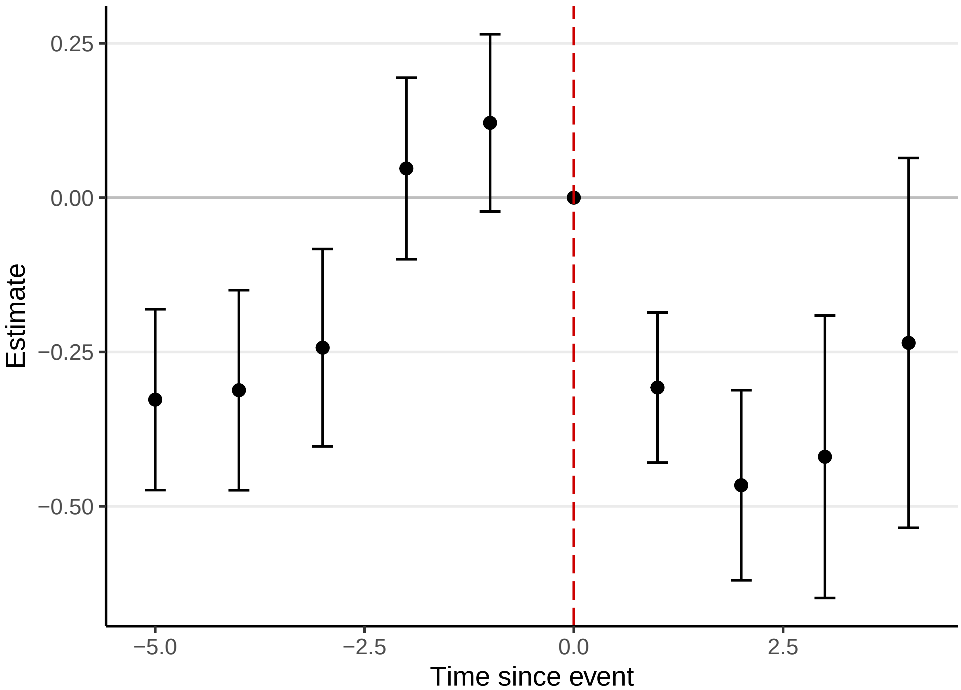


Notes: This graph shows the distribution of predicted hazards from a logistic discrete-time hazard model in equation (9). Two distributions are drawn separately for selected months by setting the peer mobilization rate to 0% and 1%, respectively. The outcome is an indicator variable, which gets value 1 for the month the individual joins the rebellion, and is 0 otherwise. The model also includes municipality fixed effects and a fifth-order polynomial of survival duration.

Figure A17: Event-Study Estimates of Repression on Mobilization



(a) First Rebel Casualty



(b) First Battle

Notes: The figure presents event-study estimates of β_τ from a staggered difference-in-differences design as defined in equation (10) (Callaway and Sant'Anna 2021). In Panel (a), the treatment is an indicator of the first rebel casualty, and in Panel (b) it is an indicator of the first major battle. The control group consists of not-yet-treated observations. The red vertical line indicates the month when repression began. Standard errors are clustered at municipal level. The error bars present 95% confidence intervals.

Table A1: Change in Public Opposition and Mortality in 1918: Heterogeneity

	ln War mortality		
	(1)	(2)	(3)
Change in public opposition × Post × SDP vote share	0.13***		
	(0.01)		
Change in public opposition × Post × Gini		0.06***	
		(0.02)	
Change in public opposition × Post × Workers houses			0.06***
			(0.02)
Change in public opposition × Post	0.13***	0.19***	0.16***
	(0.04)	(0.04)	(0.04)
Municipality fixed effects	✓	✓	✓
Time fixed effects	✓	✓	✓
Controls	✓	✓	✓
Observations	2,490	2,256	2,490
R^2	0.92	0.93	0.93

Notes: Triple difference-in-differences estimates from equation (4). The unit of observation is a municipality. SDP vote share is an indicator variable, which gets value 1 if the municipality had above median level of SDP support in 1916 elections, and is 0 otherwise. Gini is an indicator variable, which gets value 1 if the municipality had above median level of land gini in 1910, and is 0 otherwise. Workers house is an indicator variable, which gets value 1 if the municipality had a workers house in 1916, and is 0 otherwise. All regressions control for municipal fixed effects, time fixed effects, as well as baseline controls for SDP vote share, historical mobilization, log population and longitude and latitude, except the regression in column (2), which controls forhistorical mobilization, log population and longitude and latitude. Standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2: Russian Telegraph

	News on Russian Revolution	Change in Public Opposition	ln War mortality
	(1)	(2)	(3)
Telegraph	0.07**	0.30**	
	(0.04)	(0.13)	
Telegraph × Post			0.17*
			(0.09)
Municipality fixed effects			✓
Time fixed effects			✓
Controls	✓	✓	✓
Observations	415	415	2,874
R^2	0.40	0.21	0.92

Notes: The unit of observation is a municipality. Telegraph is an indicator variable, which gets value 1 if the municipality has a Russian telegraph station, and is 0 otherwise. Standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Random Sample of Top-Scoring Motives by Category

Source text	**Translation**
Panel A: Conformity motive	
"Muiden kehoituksesta."	*At the encouragement of others.*
"Pakolla viety Lintulen sahatta."	*Was taken by force from Lintula sawmill.*
"Toisten pahoituksesta."	*Due to others' influence.*
'Pakotettiin mukaan."	*Was forced to join.*
"Pelkäsi joutuvansa painastukaan alosaksi."	*Feared being pressured.*
Panel B: Material motive	
"Kun ei ollut töitä."	*Because there was no work.*
"Työpuutteen takia kätäaputyöt kun lakkautelliin p 1917 vuoden lopulla."	*Due to lack of work after relief work was abolished at the end of 1917.*
"Rullen puutteen takia ja kun pientä vaatinusta."	*Due to a lack of [unreadable] and because of minor pressure.*
"Liittyi työn puutteesta."	*Joined due to lack of work.*
"Työtösmyyden takia."	*Due to unemployment.*
Panel C: Ideological motive	
"Vapaehtoisesti."	*Voluntarily.*
'Vapaaehtoisesti anion vuoksi."	*Voluntarily for the pay.*
"Omasta tahdostaan."	*Out of own will.*
"Vapaa ehtoisesti ajan hengen mukaisest."	*Voluntarily, in accordance with the spirit of the times.*
"Vapaaehtoisesti"	*Voluntarily.*

Notes: This table shows a random sample of top-scoring motives by category. Top-scoring motives have a prediction probability among the top quantile in the respective class. The sample includes only responses with at most 200 characters.

Table A4: Change in Public Opposition and Individual Mobilization Motives in 1918

	Conformity motive	Material motive	Ideological motive	Motiveless motive
	(1)	(2)	(3)	(4)
Change in public opposition	-0.007	0.01*	0.0004	-0.004
	(0.006)	(0.006)	(0.003)	(0.005)
Controls	✓	✓	✓	✓
Observations	32,578	32,578	32,578	32,578
Mean Y	0.42	0.19	0.12	0.27
R^2	0.00	0.02	0.00	0.01

Notes: The unit of observation is an individual. Ideological motive is an indicator variable, which gets value one if the person's motive was categorized as ideological with a binary text classification model. Outcomes in columns (2) to (4) are defined similarly. All regressions include baseline controls for SDP vote share, historical mobilization, log population and longitude and latitude. Standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5: Change in Public Opposition and Differently Motivated Rebels in 1918

	ln Conformity rebels	ln Material rebels	ln Ideology rebels	ln Motiveless rebels
	(1)	(2)	(3)	(4)
Change in public opposition	0.16***	0.12***	0.09***	0.13***
	(0.04)	(0.03)	(0.03)	(0.03)
Controls	✓	✓	✓	✓
Observations	415	415	415	415
Mean Y	4.83	1.84	1.39	3.27
R^2	0.40	0.34	0.34	0.36

Notes: The unit of observation is a municipality. Ideological rebels is the number of ideologically motivated rebels per 1000 people +1. Outcomes in columns (2) to (4) are defined similarly. All regressions include baseline controls for SDP vote share, historical mobilization, log population and longitude and latitude. Standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6: Strategic Complementarity

	Joined rebellion			ln War mortality	
	Logistic discrete-time hazard model	Gaussian discrete-time hazard model	Linear probability model	OLS	
	(1)	(2)	(3)	(4)	(5)
Peer mobilization rate	0.98***	0.50***	0.06***		
	(0.17)	(0.10)	(0.02)		
Change in public opposition × Post × Subscribers				0.03**	
				(0.01)	
Change in public opposition × Post × Strikers					0.05***
					(0.01)
Change in public opposition × Post				0.22***	0.17***
				(0.04)	(0.04)
Average partial effect	0.048	0.0451	0.0613		
Observations	218,072	218,072	218,072	2,493	2,493
Mean Y	0.11	0.11	0.11	2.81	2.81
R^2	NA	NA	0.49	0.93	0.93
Fifth-order polynomial of duration	✓	✓			
Municipality fixed effects	✓	✓	✓	✓	✓
Time fixed effects			✓	✓	✓
Controls				✓	✓

Notes: The unit of observation in columns (1) to (3) is an individual, and in columns (4) and (5) a municipality. In columns (1) to (3) the dependent variable is an indicator, which gets value 1 for the month the individual joins the rebellion. In columns (4) and (5), the dependent variable is log of war mortality +1. Peer mobilization rate is the local leave-one-out mobilization rate, expressed in percentages. Subscribers is an indicator variable, which gets value 1 if the municipality has above median number of newspaper subscribers per capita, and is 0 otherwise. Strikers is an indicator variable, which gets value 1 if the municipality had above median number of strikers per capita in 1917, and is 0 otherwise. Columns (4) to (5) express triple-differences estimates from equation (4). All regressions control for municipal fixed effects. Columns (1) to (2) additionally control for a fifth-order polynomial of survival duration. Columns (3) to (5) control for time fixed effects. Columns (4) to (5) further include baseline controls for SDP vote share, historical mobilization, log population and longitude and latitude. Standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A7: Repression

	ln Rebels per capita			
	(1)	(2)	(3)	(4)
Lag of ln Rebel mortality	-0.50***	-0.58***		
	(0.03)	(0.03)		
Lag of ln Rebel mortality × Above median change in public opposition		0.06		
		(0.05)		
Lag of First battle			-0.36***	-0.44***
			(0.04)	(0.06)
Lag of First battle × Above median change in public opposition				0.09
				(0.07)
Municipality fixed effects	✓	✓	✓	✓
Time fixed effects	✓	✓	✓	✓
Observations	2,534	2,163	3,060	2,604
Mean Y	1.27	1.28	1.14	1.15
R^2	0.77	0.79	0.71	0.72

Notes: The unit of observation is a municipality. The dependent variable in all regressions is the log of rebels per capita, which measures mobilization. Lag of ln rebel mortality is the first lag of log rebel casualties per 1000 people +1. Lag of first battle is the first lag of an indicator, which gets value 1 from the first battle onwards. All regressions control for municipal fixed effects and month fixed effects. Standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A8: Change in Public Opposition and Mortality in 1918: Matched Control Group

	ln War mortality	
	Nearest neighbor	Subclassification
	(1)	(2)
Above median change in public opposition × Post	0.22***	0.36***
	(0.07)	(0.11)
Municipality fixed effects	✓	✓
Time fixed effects	✓	✓
Controls	✓	✓
Observations	2,424	2,490
R^2	0.93	0.92

Notes: Difference-in-differences estimates from equation (2), after matching on preexposure characteristics. The unit of observation is a municipality. In column (1), the matching method is nearest neighbor propensity score matching. In column (2), it is propensity score subclassification. All regressions control for municipal fixed effects, time fixed effects, as well as baseline controls for SDP vote share, historical mobilization, log population and longitude and latitude. Standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A9: Change in Public Opposition and Mortality in 1918: Alternative Specifications

	ln War mortality		
	Conley SE	Omitting Viipuri province	Empirical subscribership
	(1)	(2)	(3)
Change in public opposition × Post	0.21***	0.15***	0.32***
	(0.07)	(0.05)	(0.05)
Municipality fixed effects	✓	✓	✓
Time fixed effects	✓	✓	✓
Controls	✓	✓	✓
Observations	2,490	2,190	2,772
R^2	0.93	0.92	0.92

Notes: Difference-in-differences estimates from equation (2). The unit of observation is a municipality. In column (1), standard errors are adjusted for spatial correlation within 100km radius (Conley 1999). In column (2), Viipuri province is omitted from the sample. In column (3), the treatment is computed by using an empirical proxy of newspaper subscribers, where subscribership is based on how often a municipality is mentioned in a given journal before the abolition of censorship. All regressions control for municipal fixed effects, time fixed effects, as well as baseline controls for SDP vote share, historical mobilization, log population and longitude and latitude. In columns (2) and (3), standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A10: Change in Public Opposition and Mortality in 1918: Additional Controls

	ln War mortality						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Change in public opposition	0.19***	0.19***	0.19***	0.20***	0.21***	0.21***	0.17***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Municipality fixed effects	✓	✓	✓	✓	✓	✓	✓
Time fixed effects	✓	✓	✓	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓	✓	✓	✓
Strikers & workers houses	✓	✓	✓	✓	✓	✓	✓
Shortage		✓	✓	✓	✓	✓	✓
Swedish population, % of all			✓	✓	✓	✓	✓
Slope				✓	✓	✓	✓
Land gini					✓	✓	✓
Russian troops						✓	✓
Province-specific slopes							✓
Observations	2,490	2,490	2,490	2,484	2,256	2,256	2,256
R^2	0.93	0.93	0.93	0.93	0.93	0.93	0.94

Notes: Difference-in-differences estimates from equation (2). The unit of observation is a municipality. Strikers is the log number of strikers in 1910–1914. Workers houses is the number of working class facilities within the municipality. Shortage is a categorical variable measuring the severity of grain shortage in 1917, where 0 refers to no shortage and 2 refers to serious shortage (Rantatupa 1979). Slope is the log median terrain slope class in the municipality. Russian troops is the log total of Russian troops stationed in the municipality in 1917. All regressions control for municipal fixed effects, time fixed effects, as well as baseline controls for SDP vote share, historical mobilization, log population and longitude and latitude. Standard errors are clustered at municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Digitization of the Interrogation Records

To study the mobilization motives of the Finnish Civil War, I utilize 750,000 treason court documents scanned and published by the National Archives of Finland. I digitize a subsample of the documents, specifically interrogation records, using a tailored multistep handwritten text recognition (HTR) pipeline. The pipeline was run sequentially on several GPUs rented via Google Colab. Steps 1 to 3 were primarily executed on T4 GPUs, while text recognition in step 4 was performed using an A100 GPU for enhanced processing speed. A sketch of the pipeline is shown in Figure B1.[6]

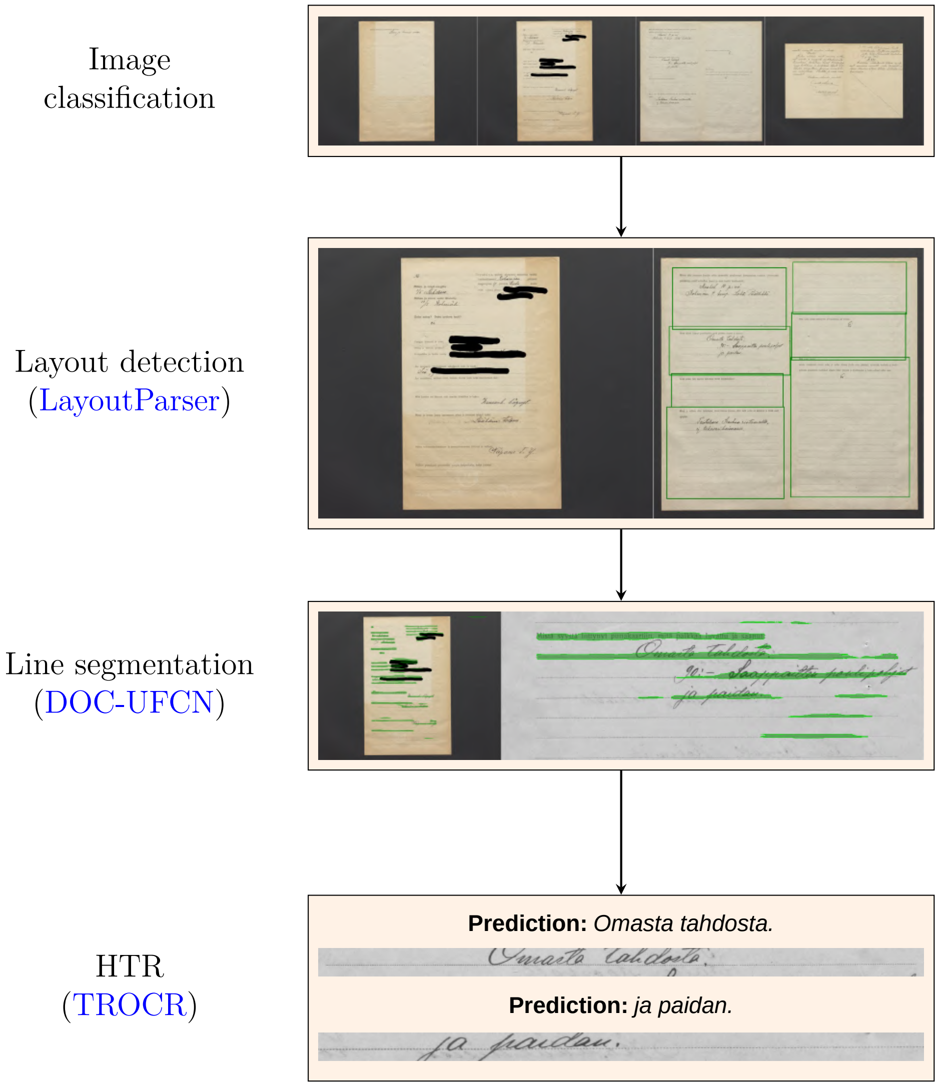
First step of the pipeline is to identify the interrogation records among the heterogeneous set of treason court documents. This would be a tedious task to carry out manually, as for any given defendant there is a diverse set of court documents (roughly ten per person) in irregular order. Interrogation records consist of two separate pages, one including the personal information of the defendant, and the other regarding questions and responses about their motives and actions during the conflict (see Figure B2). I refer to these two pages as the *id page* and the *motive page*, respectively. In terms of a classification model, the pages represent two distinct classes. Using a pre-trained Vision Transformer (ViT) model, I fine-tune an image classification model to categorize the court records into id pages and motive pages. Once the fine-tuning is finished, the classification model reaches validation loss of 0.002. By applying the fine-tuned classification model to draw predictions for all treason court documents, I am left with a sample of around 150,000 interrogation records.

The second step in the pipeline is to detect and crop response boxes within the standardized layout of the interrogation record forms. To achieve this, a layout detection model is trained using LayoutParser, an open-source Python library for deep learning based document image analysis (Shen et al. 2021). I employ pre-trained Mask R-CNN R-50-FPN 3x model as a baseline, and fine-tune it using a sample of pages where the layout detection is done by hand. To create the training data, I use Label Studio, which is a tool designed for annotating datasets for machine learning tasks. Notably, layout detection is only applied to the motive pages, as line segmentation in the next phase of the pipeline is sufficient for HTR process for the id pages.

In the third step of the pipeline, the id pages and the cropped response boxes from the motive pages are segmented into separate lines of text. The line segmentation is carried out using another deep learning based tool, Doc-UFCN (Boillet, Kermorvant, and Paquet 2021, 2022), a fully convolutional neural network for text line

[6] For examples of similar pipelines, see Stenhammar (2023) and Dahl et al. (2023).

Figure B1: HTR Pipeline



Notes: This figure shows a sketch of the HTR pipeline applied to digitize interrogation records. The interrogation records were a part of the background material which treason courts relied on when convicting the accused rebels. The first step of the pipeline is to identify the relevant records using an image classification model. The second step is to detect the layout of the records to crop each answer box or a cell corresponding to a particular question. The third step is to segment the record pages into lines of text. The fourth step is to use a transformer-based handwritten text recognition model to predict the text in each line.

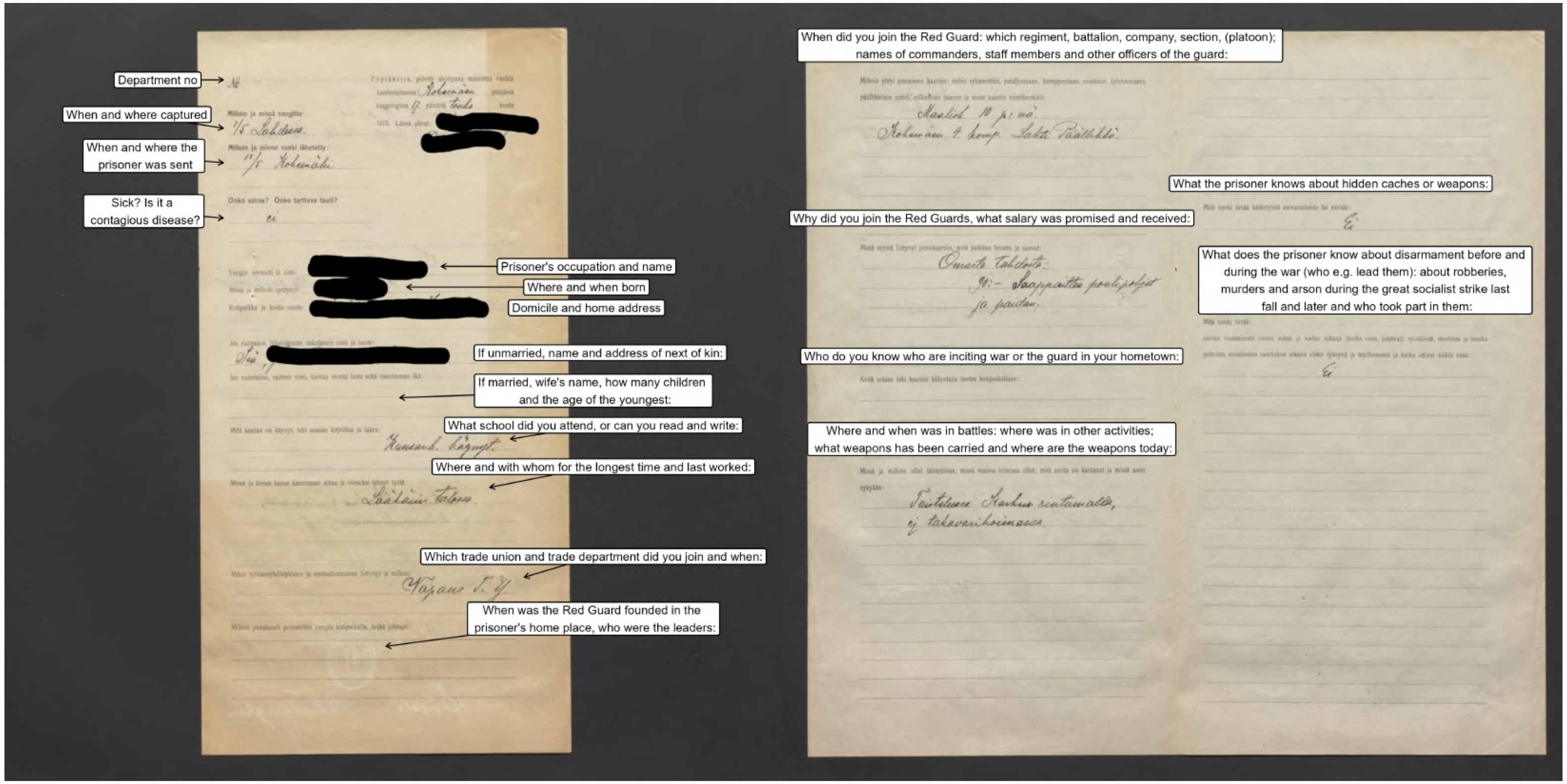
detection. I use Doc-UFCN's generic historical line detection model pre-trained on ten historical document datasets and made available by Hugging Face. By deploying this model, each id page and response box is segmented into individual lines of handwritten text for subsequent recognition. The segmentation step is crucial, as current state-of-the-art handwritten recognition models are largely built to work with single text-line inputs. Upon completing the line segmentation, approximately 3.5 million text-line images remain for recognition.

The fourth and final step of the pipeline is text recognition. For this task, I employ TrOCR (Li et al. 2023). TrOCR is an innovative text recognition model, as it leverages the powerful Transformer architecture (Vaswani 2017) twice: first when encoding the images into numbers, and second when decoding these numbers into text. This is particularly useful for text recognition, as TrOCR is then able to harness the capacity of large language models, proven to demonstrate deduction skills which closely resemble those of human (Horton 2023; Korinek 2023).

I utilize a TrOCR model provided by the Swedish National Archives (Riksarkivet), which has been fine-tuned on handwritten Swedish historical documents. I perform second-stage fine-tuning using a sample of my own training data produced in the previous step of the pipeline, read and labeled by the author. The final model achieves character error rate (CER) of 5%, meaning that the model is able correctly predict 95% of letters in an unseen validation dataset. An example of the text recognition step is shown in Figure B3.

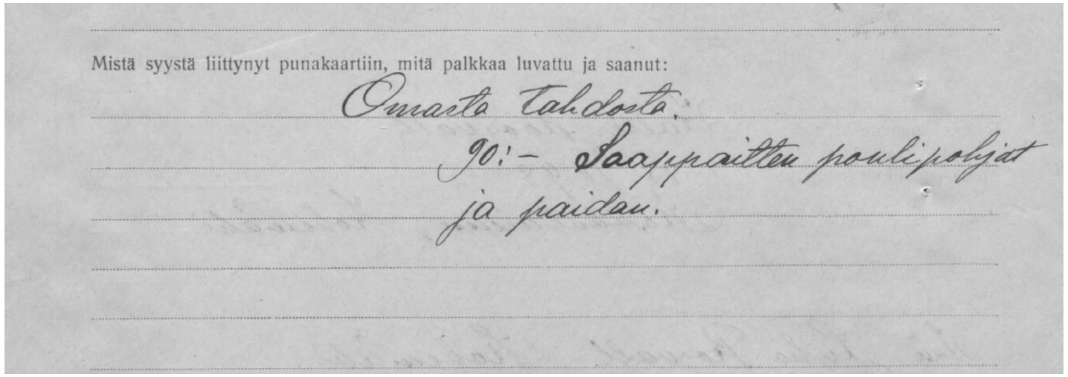
The interrogation records provide information e.g. on the accused's occupation, name, birthplace, domicile, which Red Guard department they belonged to, timing of joining the Guard and most importantly, motive for enlisting in the Red Guard, in their own words. I employ a number of postprocessing measures to turn the transcribed text into a dataset where the unit of observation is a single defendant. To start with, I utilize fuzzy matching and sentence embeddings to identify questions of interest in the text data. Sentence embeddings are created using a multilingual SentenceTransformer (Reimers and Gurevych 2019, 2020). The identification is executed according to the following ruleset:

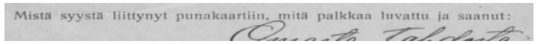
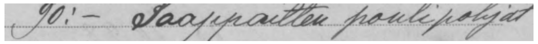
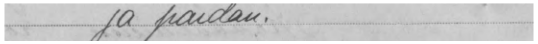
Figure B2: Excerpt of Interrogation Records



Notes: This figure shows an excerpt of the interrogation records. Personal information is censored retrospectively by the author.

Figure B3: Excerpt of Accused's Motive


Mistä syystä liittynyt punakaartiin, mitä palkkaa luvattu ja saanut:
Omasta tahdosta.
90:- Saappaitten poulipohjat
ja paidan.

Segmented line	Prediction	Translation
	Mistä syystä liittynyt punakaartiin, mitä palkkaa luvattu ja saanut:	*Why did you join the Red Guards, what salary was promised and received:*
	Omasta tahdosta.	*Of my own free will.*
	90: - Saappaitten poulipohjat	*90: - Half soles of boots.*
	ja paidan.	*and a shirt.*

1. If text-line i fuzzy matches a question within the complete list of questions with 90% confidence, classify the line as a question

2. If text-line i is 90% similar to a question within the complete list of questions in terms of cosine similarity, classify the line as a question

Having identified questions in the text data, I drop nonreadable documents by sparing only pages with at least one identified question.

I restrict the sample of interrogation records to include only the documents where the motive page follows right after the id page in the scanned archival material. The restriction is to ensure that each motive page is paired with the correct id page, and it leaves me with 98.9% of the sample.

As the information on the accused's homeplace suffers from missing observations, typological errors and non-standard reporting (sometimes, for instance, including only the accused's home address), I implement a hierarchical approach to match each person to a given treatment value, which varies at municipal level:

1. I scrape a crowdsourced subset of the accused's names and hometowns, assembled from interrogation records by a range of archival users, notably genealogists. This data is hosted by Digihakemisto.[7]

2. I extract names of the accused and their homeplaces by using named-entity recognition (NER) on said crowdsourced dataset, along with the variables describing the accused's name, domicile, and the location of the Guard in which the defendant held membership. I perform NER using FinBERT (Virtanen et al. 2019; Luoma et al. 2020), a Transformer-based language model which has been pre-trained on both historical and contemporary Finnish text data.[8]

3. I lemmatize the extracted municipality names by applying Stanza, a Python package for natural language processing which has been initialized by a large number of different languages, including Finnish (Qi et al. 2020). Lemmatization reduces the extracted munipality names to their base form, improving the quality of matches.

4. After the preceding steps, I have consistent but patchy information on each defendant's homeplace. I fuzzy match this data to a complete list of Finnish municipalities leveraging several sources of information. The matching logic is as follows

[7] https://digihakemisto.net/

[8] FinBERT was made available by Finnish National Archives (Kansallisarkisto) via Hugging Face.

(a) Match if accused i's domicile has a 90% confident match within the complete list of Finnish municipalities

(b) Match if accused i's location entity in the crowdsourced dataset has a 90% confident match within the complete list of Finnish municipalities AND a person entity in the crowdsourced dataset matches the handwritten name with 60% confidence

(c) Match if the location of accused i's Guard department has a 90% confident match within the complete list of Finnish municipalities

(d) Match if accused i's handwritten name has a 95% confident match within a complete crosswalk of civil war prisoners, casualties and their homeplaces

(e) If accused i's domicile municipality is still missing, settle with the closest match in the complete list of Finnish municipalities

Using the five-step process, I am able to match 83% of the people in the data to a municipality.

C Validation of the Treatment Variable

This section lays out the arguments for using a particular dictionary of inflammatory words to measure public opposition.

Gentzkow, Kelly, and Taddy (2019) summarize the outline of text analysis in three simple steps:

1. Represent raw text $\mathcal{D}$ as a *document-token matrix* $\boldsymbol{C}$
2. Map $\boldsymbol{C}_{n\times p}$ to predicted values $\hat{\boldsymbol{V}}_{n\times k}$ of unknown outcomes $\boldsymbol{V}_{n\times k}$, and
3. Use $\hat{\boldsymbol{V}}$ for subsequent analysis

In this paper, my objective is to proxy the extent of latent public opposition $\boldsymbol{v}_{n\times 1}$ in a set of n newspaper journals with some text-based measure $\hat{\boldsymbol{v}}_{n\times 1}$. I adapt the popular *bag-of-words* model to represent the journals. To see what this entails, I follow the notation in Ash and Hansen (2023) to denote the content of a journal (document) d as a vector

$$\boldsymbol{w}_d = (w_{d,1}, \ldots, w_{d,N_d}),$$

where $w_{d,i}$ is term i of journal d and N_d is document length, i.e. the total number of words in journal i.

Next, each unique term in the newspaper corpus vocabulary is assigned with an unique index from integers $1, \ldots, V$, where V is the total number of unique terms. The count of term v in journal d can now be defined as

$$x_{d,v} = \sum_n 1\{w_{d,n} = v\},$$

so that $\boldsymbol{x}_d = (x_{d,1}, \ldots, x_{d,V})$ is the vector of counts.

By stacking $\boldsymbol{x}_d$ across rows, we get a matrix

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_N \end{bmatrix},$$

which is known as the *document-term matrix*. In the bag-of-words model, the sparse document-term matrix $\boldsymbol{X}$ represents our document-token matrix $\boldsymbol{C}$.

To find a measure $\hat{\boldsymbol{v}}$ of the latent variable of interest, $\boldsymbol{v}$, we need a mapping

$$\boldsymbol{X} \rightarrow \hat{\boldsymbol{v}}.$$

For this purpose, I employ *keyword counting*, which is a dictionary-based method (Gentzkow, Kelly, and Taddy 2019; Ash and Hansen 2023) for mapping terms into concepts. In keyword counting, each document is converted into a count over matched terms in a concept-related dictionary $\mathfrak{D}$. Here, dictionary $\mathfrak{D}$ is a set of inflammatory words meant to capture public dissent. To make the counts comparable across journals, I normalize them by the total number of stories per journal. The predicted public opposition of journal d, $\hat{v}_d$, is then given by

$$\hat{v}_d = \frac{1}{M_d} \sum_{v \in \mathfrak{D}} x_{d,v} = \frac{z_d}{M_d}$$

where M_d is the number of news articles in journal d, and z_d is the raw count of inflammatory words in journal d.

Selecting the term set for the concept dictionary $\mathfrak{D}$ involves balancing several competing objectives, such as generalizability and contextual specificity. A common strategy in prior research is to employ external, off-the-shelf dictionaries curated by domain experts (Enke 2020). This approach is advantageous because it reduces the degrees of freedom in the researcher's decision-making process, thereby minimizing the potential for manipulation. However, it comes with the drawback of limited context awareness: it is improbable to find an external dictionary that precisely captures the concept of interest, particularly when considering specific languages, historical contexts, and text types. An alternative method is to construct the term set independently. Although this method allows for greater customization, it is also more vulnerable to researcher bias and potential manipulation.

In the context of Finnish Civil War, the pioneering work by Turunen (2021) has shed new light on the inflammatory language of 1917 in the newspapers. Turunen examines how the working class' ambiance evolved over the year in admirable detail, word by word, from a hopeful spring to a grim winter.

I rely on the historical expertise of Turunen (2021), and define the concept dictionary $\mathfrak{D}$ as the following set of terms[9]

$$\mathfrak{D} = \Big\{\text{“revolution”, “democracy”, “anarchy”, “oppression”, “freedom”, “butcher”}\Big\}.$$

Rhetoric in 1917 according to Turunen (2021) “Freedom” (*vapaus*) was among the first trending themes of spring 1917. Freedom was topical, because it was an apt word to describe the atmosphere of being liberated from under the oppressive power (*sortovalta*) of the Russian Empire (Suodenjoki and Turunen 2017). Concretely, freedom was a reference to the abolition of *“Russification acts, censorship and other restrictions on civil rights”* by the Provisional Government's manifesto on 20th of March, 1917 (Turunen 2021, p. 269). Due to newfound freedom, mood among the populace was largely optimistic: the future of Finland, although uncertain, appeared bright to most.

Optimism was soon accompanied by timely demands for expanded democracy (*demokratia*). After the political deadlock caused by the Tsar's vetoes, the Social Democrats, in particular, believed that society could finally be improved through parliamentary means. In April, morale was high as one of the socialists' long-standing goals – an eight-hour workday – was enacted into law. Over time, the democracy discourse started to absorb new tones. The political upheaval in Russia came to be described as “revolutionary democracy”, signaling a shift in power away from imperialists and supporters of the old regime.

According to Turunen, the summer of 1917 represented a new “watershed” in public rhetoric, *“the crucial moment when the high hopes of the spring changed into the anticlimax of the autumn”* (Turunen 2021, p. 307). In July 1917, Russian Provisional Government dissolved Finnish parliament, where social democrats presented the majority in bourgeois coup d'état (Alapuro 2018). *“In the socialist understanding, the bourgeois parties had betrayed true democracy”* (Turunen 2021, p. 279). Along imperial oppressors (*sortajat*), newspapers began to print more and more references to bourgeois or capitalist oppressors. Discontent resulted in strikes and rioting, which the bourgeois newspapers described as “anarchy” (*anarkia*). Meanwhile, socialist editors argued that the lack of municipal democracy and food shortages fueled local anarchy. To aggravate class antagonism to the edge, the bourgeois government was referred to as “legal anarchy”.

By autumn 1917, Finnish society began to slide toward chaos. A verbal manifestation of this trajectory was growing use of the word “butcher” (*lahtari*), a

[9] Taking into account spelling, stemming and bilingual source text, the full set of keywords include: *vallankum*, wallankum*, revolution, Demokr*, demokr*, anark*, sorto, förtryck, vapau*, frihet, lahtari, slaktare.*

particularly pejorative nickname for the bourgeoisie, which *"conveyed the idea of cruel and inhumane killings of other people"* (Turunen 2021, p. 292). On top of the inspiring news coverage of the Russian revolution, "working-class revolution" (*työväenluokan vallankumous*) became a recurring talking point in the socialist press. Eventually, revolution rhetoric acted as public threat toward the bourgeoisie. For instance, as part of the general strike in November 1917, social democratic party leaders published the so called "We Demand" program, which stated that *"throughout this revolutionary period, it [the Finnish bourgeoisie] has not made any honest concession to the working people in its politics, as if it had completely lost its ability to understand that a tightly drawn bow snaps if one does not release it in time"* (Turunen 2021, p. 289).

Manual audit To make sure that keyword counting is a working approach to measure public discontent, I follow Baker, Bloom, and Davis (2016), Gentzkow, Kelly, and Taddy (2019), Gennaro and Ash (2022), and Lippmann (2022) by conducting a manual audit for a subsample of the news articles. Specifically, I draw a sample of 400 news articles from the data, and assign the article as "inflammatory" if it contains at least one inflammatory word. I then inspect these articles myself, and evaluate whether they actually align with public opposition. To deal with the fact that only 1.8% of the articles in the whole dataset have a positive keyword count, I use random undersampling on the majority class of neutral news articles.

Table C1: Performance Metrics for Different Models

Metric	**Whole Sample**	**Before the Abolition**	**After the Abolition**
Precision	0.51	0.17	0.85
Recall	0.94	0.94	0.94
False negative rate	0.06	0.06	0.06
F1 Score	0.66	0.29	0.89

Notes: Performance metrics based on a random sample of 400 news articles. The majority class of neutral news was drawn using random undersampling. Precision is the ratio of true positive predictions to the total predicted positives. Recall is the ratio of true positive predictions to all actual positives. The false negative rate is the ratio of false negatives to the total actual positives. The F1 score is the harmonic mean of precision and recall.

Tables C1 and C2 present performance metrics and confusion matrices for the whole sample, and split before and after abolition of censorship, respectively. The

Table C2: Manual Audit Confusion Matrices

	Predicted negative	Predicted positive
Negative	194	98
Positive	6	102

(a) Whole Sample

	Predicted negative	Predicted positive
Negative	99	83
Positive	1	17

(b) Before the Abolition

	Predicted negative	Predicted positive
Negative	95	15
Positive	5	85

(c) After the Abolition

metrics indicate, that the selected keywords do a good job in measuring revolutionary writing after the abolition of censorship. In this period, out of all the articles categorized as inflammatory, 85 percent reflect public discontent according to human judgement. Moreover, the keywords are able to capture 94 percent of all inflammatory news in the random sample. Complementarily, 6% of revolutionary texts go unobserved. Before the abolition of censorship, the keywords catch a lot of noise: only 17% of all the articles categorized as inflammatory are actually revolutionary. This is expected, since during censorship public dissent was effectively banned: just 0.5% of the articles had then a positive keyword count, compared to 3.5% once censorship was lifted.

Reassuringly, keyword counting misses only a negligible fraction of revolutionary news in a manual audit. However, low precision before the abolition of censorship introduces random measurement error, implying that the coefficient of interest is biased toward zero.

Placebo treatment To reassure that the correlation between $\Delta PublicOpposition_m$ and war mortality is unrelated to the way the treatment is computed, I run a placebo exercise. I construct placebo treatment using the three most common stopwords in the Finnish language, namely "and", "which" and "that", and calculate the change in the usage of the stopwords following equation (1) as a shift-share measure:

$$\Delta Placebo_m = \sum_d w_{md} \Delta Stopwords_d.$$

I then re-estimate the event-study specification (3) using $\Delta Placebo_m$ as the treatment.

The event-study estimates are reported in Figure C1. Curiously, the estimates collectively suggest that an increase in the usage of stopwords has a negative effect on war mortality, which is statistically significant at the 5% level. However, with respect to overall mortality, no clear effect can be found. A potential explanation for this observation is that the neutral language of stopwords to some extent crowds out sentiment-carrying inflammatory words. Reassuringly, their effect on mobilization is, if anything, negative, mitigating the worry of a mechanical correlation between public opposition and mortality.

Adding and subtracting words from $\mathfrak{D}$ Specifying a particular dictionary to measure public opposition is potentially vulnerable to the choice of words. I show that my measure of public opposition is robust to small perturbations in the applied dictionary by shrinking and expanding the set of words using machine judgement (Gennaro and Ash 2022; Ash and Hansen 2023). First, I construct word embeddings for my text data with word2vec algorithm (Mikolov et al. 2013) applied to the full corpus of news published after the abolition of censorship. Next, I find words which are most similar to the average embedding vector of dictionary $\mathfrak{D}$ based on cosine similarity. The twenty most similar words are presented in Table C3.

I perturb $\mathfrak{D}$ in three different ways. First, I expand the dictionary by 50% by adding three most similar new words to the base dictionary's average word embedding. From table C3, one can see that these three words are "reaction", "tyranny" and "violence". Second, I shrink the dictionary by 50% by removing the three least similar words to the base dictionary's average word embedding. This leaves me with a reduced dictionary with the terms "oppression", "revolution" and "anarchy". Third, I study whether the results hinge on any one word in particular by constructing six leave-one-out dictionaries with one base word excluded at a time.

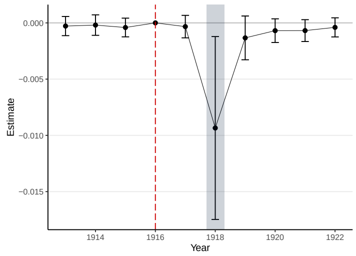
I then re-estimate the event-study regression (3) for each perturbation separately. The results are shown in Figures C1 and C2. The pattern of estimates confirms, that the relationship between public opposition and mobilization is robust to small perturbations regarding how the concept of interest is measured.

Table C3: Twenty Most Similar Words to Dictionary $\mathfrak{D}$

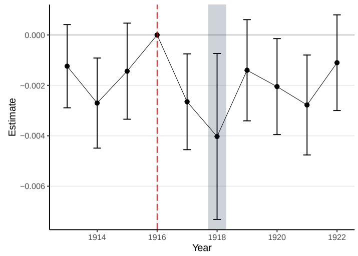
Source word	**Translation**	**Cosine similarity**
Sorto	*Oppression*	0.75
Vallankumous	*Revolution*	0.75
Taantumus	*Reaction*	0.75
Anarkia	*Anarchy*	0.74
Hirmuwalta	*Tyranny*	0.72
Väkivalta	*Violence*	0.72
Wallankumous	*Revolution*	0.72
Sosialismi	*Socialism*	0.72
Hallitusjärjestelmä	*Regime*	0.71
Kansanwalta	*Democracy*	0.71
Sosialidemokratia	*Social democracy*	0.71
Surkeus	*Misery*	0.71
Itsevaltius	*Despotism*	0.71
Vapaus	*Freedom*	0.71
Kumous	*Upheaval*	0.70
Mahti	*Might*	0.70
Yhteiskunta	*Society*	0.70
Sortojärjestelmä	*System of oppression*	0.70
Köyhälistö	*The poor*	0.70
Sekasorto	*Chaos*	0.70

Notes: This table lists the twenty most similar words to the average word vector of dictionary $\mathfrak{D}$ in the source text. The source text vocabulary was vectorized using word2vec. Recurring words with different spelling are highlighted in blue.

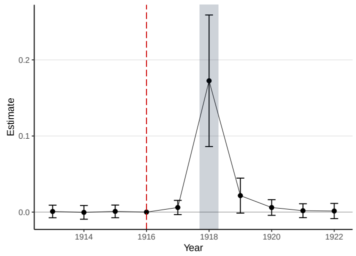
Figure C1: Event-Study Estimates of a Change in Public Opposition on War Mortality and Overall Mortality: Different Dictionaries



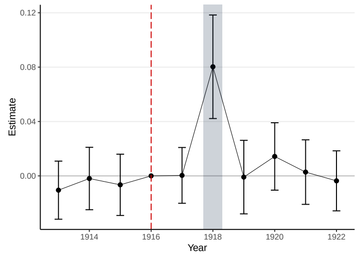
(a) Stopwords and War Mortality



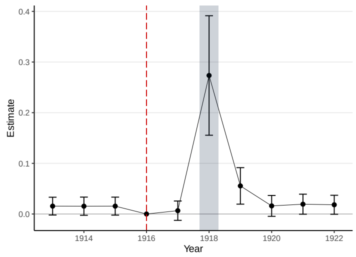
(b) Stopwords and Overall Mortality



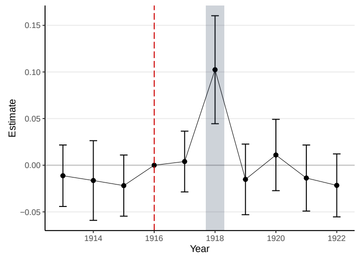
(c) Large Dictionary and War Mortality



(d) Large Dictionary and Overall Mortality



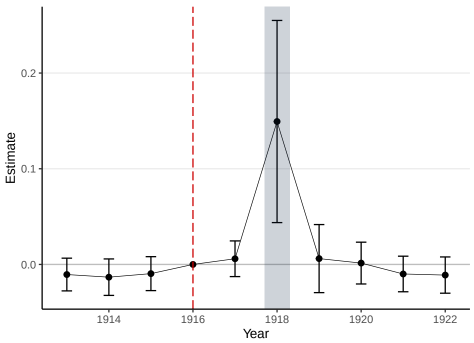
(e) Small Dictionary and War Mortality



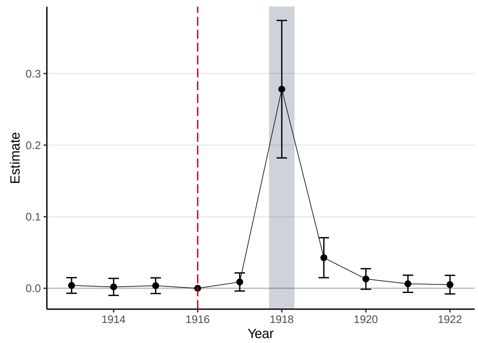
(f) Small Dictionary and Overall Mortality

Notes: **Panels (a)-(b)**: The figures show event-study estimates of a change in stopwords ("and", "which" and "that") on war mortality and overall mortality, respectively. **Panel (c)-(d)**: The figures show event-study estimates of a change in expanded dictionary ("revolution", "democracy", "anarchy", "oppression", "freedom", "butcher", "reaction", "tyranny" and "violence") on war mortality and overall mortality, respectively. **Panel (e)-(f)**: The figures show event-study estimates of a change in reduced dictionary ("revolution", "anarchy", and "oppression") on war mortality and overall mortality, respectively. The unit of observation is a municipality. All models include municipality and year fixed effects, as well as baseline controls for SDP vote share, historical mobilization, log population in 1916, longitude and latitude, each interacted with year effects. The red vertical line indicates the year preceding the abolition of censorship. Standard errors are clustered at municipal level.

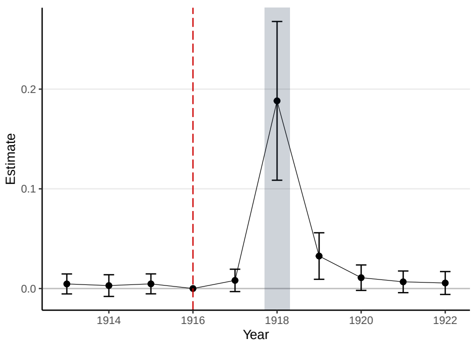
Figure C2: Event-Study Estimates of a Change in Public Opposition on War Mortality: Leave-One-Out Dictionaries



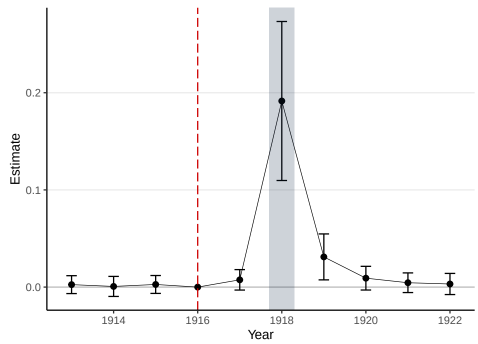
(a) Excluded Word: Revolution



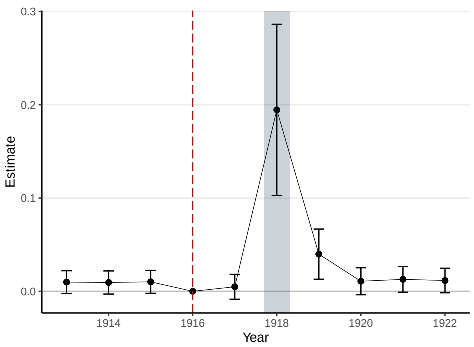
(b) Excluded Word: Democracy



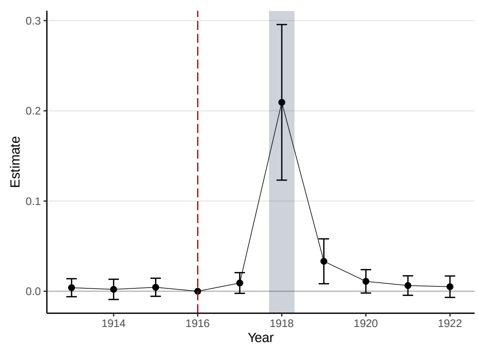
(c) Excluded Word: Anarchy



(d) Excluded Word: Oppression



(e) Excluded Word: Freedom



(f) Excluded Word: *Butcher*

Notes: Each sub-figure presents event-study estimates of β_τ from equation (3), while excluding one base word at a time. The unit of observation is a municipality. All models include municipality and year fixed effects, as well as baseline controls for SDP vote share, historical mobilization, log population in 1916, longitude and latitude, each interacted with year effects. The red vertical line indicates the year preceding the abolition of censorship. Standard errors are clustered at municipal level.

D Validation of the Outcome Variable

This section justifies the use of war mortality as a proxy for mobilization.

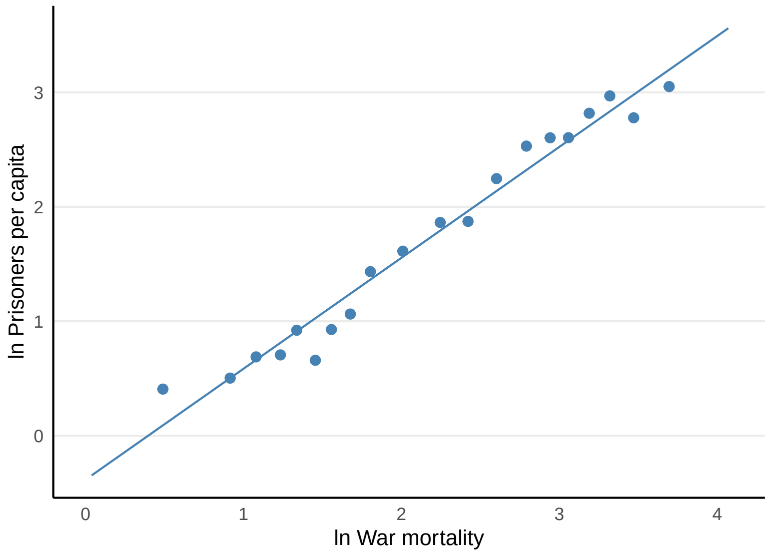
I use war mortality as defined in WarVictimSampo 1914–1922 as my main measure of local mobilization. As described in Section 3, WarVictimSampo documents all war-related deaths in Finland from 1914 to 1922, the bulk of which consist of insurgents who lost their lives in the Finnish Civil War. By definition, war mortality thus only measures dead rebels. The benefit of using war mortality as an outcome is twofold: for one, the WarVictimSampo reports war-related deaths both before and after the civil war, allowing a local difference-in-differences setup. Second, the data contains granular domicile information at village-level, which secures enough observations for a spatial regression discontinuity design. Therefore, both my identification strategies hinge on the unique features of the war mortality data. War mortality does not capture all rebels, however, as many of them outlive the failed rebellion.

For the purposes of this study, I construct two new alternative measures of mobilization to compare war mortality with. First, I scrape a sample of 30,000 captives from *Sotavankilaitoksen vankikortisto*, a dataset digitized and hosted by SukuHaku.[10] This sample consists of all the captives whose surnames start with a letter ranging from *A* to *L*. Apart from the geographical clustering of certain family names, I argue that the sample provides a fairly representative measure of the insurgents' spatial distribution. Second, I compile another proxy of the sum of local revolutionaries using the interrogation records. This data includes the universe of rebels who were imprisoned and interrogated since late May 1918.

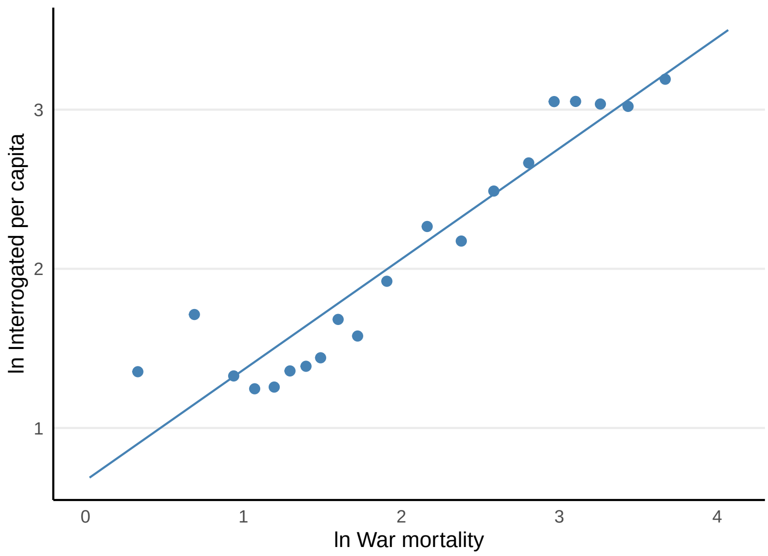
Figure D1 presents the binscatters from bivariate regressions of both proxies on war mortality. The plots illustrate, that war mortality is tightly correlated with the sample of captives and the interrogated prisoners, collected from separate sources.

[10] https://sukuhaku.genealogia.fi/.

Figure D1: Relationship Between War Mortality and Other Measures of Mobilization



(a) Prisoners Per Capita



(b) Interrogated Per Capita

Notes: **Panel (a)**: The figure shows binscatter of war mortality and prisoners per capita in a sample of prisoners from *Sotavankilaitoksen vankikortisto*. The sample consists of all the captives whose surnames start with a letter ranging from *A* to *L*. The unit of observation is municipality. **Panel (b)**: The figure shows binscatter of war mortality and the number of interrogated per person from the universe of interrogated insurgents in treason court archives. The unit of observation is municipality.

E Model Appendix

Each citizen decides whether to attack the regime

$$a_i = \begin{cases} 1 & \text{if } i \text{ attacks} \\ 0 & \text{if } i \text{ does not attack} \end{cases}$$

The regime is overthrown if

$$\theta < S$$

where θ represents government strength, and is private information to the regime. S is aggregate attack on the regime.

Citizens receive two kinds of signals regarding θ. First, newspapers send a noisy public signal $E[\theta] = y$, where θ follows a normal distribution

$$\theta \sim N\left(y, \frac{1}{\sigma_\theta}\right)$$

Second, each citizen i gets a private signal, x_i, which consists of e.g. interactions with family, friends and neighbors

$$\begin{aligned} & x_i = \theta + \varepsilon_i \\ & \varepsilon_i \sim N\left(0, \frac{1}{\sigma_\varepsilon}\right), \quad \varepsilon_i \text{ i.i.d.}, \quad \theta \perp\!\!\!\perp \varepsilon_i \end{aligned}$$

Citizens also receive another noisy public signal of payoff for joining the cause, $E[b] = z$, which the newspapers are able to manipulate. b is assumed to follow a uniform distribution

$$b \sim U[0, 2z]$$

Citizen i's payoff structure is

		$S > \theta$	$S \leq \theta$
Player i	$a_i = 1$	$E[b]$	$E[b] - c$
	$a_i = 0$	0	0

which demonstrates that joining the rebellion gives an ideological or reputational payoff, $E[b]$.

Regime survival condition Consider a cutoff equilibria, in which player j decides to join a revolution if he gets and indication that the regime is sufficiently weak, $x \leq \hat{x}$

$$a_j = 1 \iff x_j \leq \hat{x}$$

The total share of rebels then turns out to be

$$\begin{aligned} S(\theta) &= P(x_i \leq \hat{x} \mid \theta) \\ &= P(\theta + \varepsilon_i \leq \hat{x} \mid \theta) \\ &= F_{\varepsilon_i}(\hat{x} - \theta) \\ &= F_{\sqrt{\sigma_\varepsilon}\varepsilon_i}(\sqrt{\sigma_\varepsilon}(\hat{x} - \theta)) \\ &= \Phi(\sqrt{\sigma_\varepsilon}(\hat{x} - \theta)) \end{aligned}$$

The regime is replaced if and only if $S(\theta) > \theta$, where regime threshold, $\hat{\theta}$, solves for the regime survival condition

$$S(\hat{\theta}) = \hat{\theta} \iff \Phi(\sqrt{\sigma_\varepsilon}(\hat{x} - \hat{\theta})) = \hat{\theta} \tag{11}$$

Citizen indifference condition Player i finds it optimal to revolt if

$$\begin{aligned} & u(a_i = 1) \geq u(a_i = 0) \\ & \rho E[b] + (1 - \rho)(E[b] - c) > 0 \\ & \rho z + z - c - \rho z + \rho c > 0 \\ & z > (1 - \rho)c \end{aligned}$$

where $\rho = P(S(\theta) > \theta \mid x_i)$ is citizen's posterior belief of regime change, given signal x_i.

In the previous section, we established that regime change hinges on the regime threshold

$$S(\theta) > \theta \iff \theta \leq \hat{\theta}$$

Thus, the belief $P(S(\theta) > \theta \mid x_i)$ can be otherwise expressed as

$$P(S(\theta) > \theta \mid x_i) = P(\theta \leq \hat{\theta} \mid x_i)$$

To derive $P(\theta \leq \hat{\theta} \mid x_i)$, we need to find the (unnormalized) posterior distribution $f(\theta \mid x_i)$, which is determined by Bayes' rule

$$f(\theta \mid x_i) \propto f(\theta) f(x_i \mid \theta)$$

Notice that the conditional density $f(x_i \mid \theta)$ is given by

$$\begin{aligned} f_{x_i|\theta}(x_i \mid \theta) &= f_{\theta+\varepsilon_i|\theta}(x_i \mid \theta) \\ &= f_{\varepsilon_i|\theta}(x_i - \theta \mid \theta) \\ &\overset{\varepsilon_i \perp\!\!\!\perp \theta}{=} f_{\varepsilon_i}(x_i - \theta) \\ &= \sqrt{\sigma_\varepsilon}\phi(\sqrt{\sigma_\varepsilon}(x_i - \theta)) \end{aligned}$$

So that by Gaussian updating, the posterior $f(\theta \mid x_i)$ then is

$$\begin{aligned} f(\theta \mid x_i) &\propto f(\theta) f(x_i \mid \theta) \\ &= \frac{1}{\sqrt{2\pi\frac{1}{\sigma_\theta}}} \exp\left\{\frac{-(\theta - y)^2}{2\frac{1}{\sigma_\theta}}\right\} \frac{1}{\sqrt{2\pi\frac{1}{\sigma_\varepsilon}}} \exp\left\{\frac{-(x_i - \theta)^2}{2\frac{1}{\sigma_\varepsilon}}\right\} \\ &= N\left(\lambda x_i + (1-\lambda)y, (\sigma_\varepsilon + \sigma_\theta)^{-1}\right) \end{aligned}$$

$P(\theta \leq \hat{\theta} \mid x_i)$ then turns out to be

$$\begin{aligned}
P(\theta \leq \hat{\theta} \mid x_i) &= F_{\theta|x_i}(\hat{\theta} \mid x_i) \\
&= F_{\theta-\lambda x_i-(1-\lambda)y|x_i}(\hat{\theta} - \lambda x_i - (1-\lambda)y) \\
&= F_{\sqrt{\sigma_\varepsilon+\sigma_\theta}(\theta-\lambda x_i-(1-\lambda)y)|x_i}(\sqrt{\sigma_\varepsilon + \sigma_\theta}(\hat{\theta} - \lambda x_i - (1-\lambda)y)) \\
&= \Phi\left(-\sqrt{\sigma_\varepsilon + \sigma_\theta}(\lambda x_i + (1-\lambda)y - \hat{\theta})\right) \\
&= 1 - \Phi\left(\sqrt{\sigma_\varepsilon + \sigma_\theta}(\lambda x_i + (1-\lambda)y - \hat{\theta})\right)
\end{aligned}$$

Citizen i thus finds it optimal to attack if and only if $z > \Phi\left(\sqrt{\sigma_\varepsilon + \sigma_\theta}(\lambda x_i + (1-\lambda)y - \hat{\theta})\right) c$, where participation threshold $\hat{x}$ solves for the citizen indifference condition

$$z = \Phi\left(\sqrt{\sigma_\varepsilon + \sigma_\theta}(\lambda \hat{x} + (1-\lambda)y - \hat{\theta})\right) c \tag{12}$$

Newspaper editor's problem Newspaper editor k's optimization problem is

$$\max\left\{0, \quad \text{argmin}_{y>0,z>0}\ (y_k - y)^2 + (z_k - z)^2 + p(z-y)\right\}$$

Editor k wants to set the two signals y and z as close as possible to their respective bliss points, while facing a penalty p for anti-government slant, $-y$, and agitative writing, z. The FOCs are

$$\begin{aligned}
&-2(y_k - y) - p = 0 \\
&-2(z_k - z) + p = 0 \\
&\begin{cases} y = y_k + \frac{p}{2} \\ z = z_k - \frac{p}{2} \end{cases}
\end{aligned} \tag{13}$$

where we additionally assume that

$$z_k - \frac{p}{2} - c < 0 < z_k - \frac{p}{2}$$

This assumption simply ensures that joining a successful revolution is believed to be worthwhile, while joining a failing revolution is more damaging than standing out.

Equilibrium The two equilibrium conditions (11) and (12) and the FOCs from the editor's problem form a system of four equations with four unknowns. We can start solving this system by finding an expression for $\hat{x}$ from the regime survival condition

$$\begin{aligned}
\sqrt{\sigma_\varepsilon}(\hat{x}-\hat{\theta}) &= \Phi^{-1}(\hat{\theta}) \\
\hat{x} &= \frac{\Phi^{-1}(\hat{\theta})}{\sqrt{\sigma_\varepsilon}} + \hat{\theta}
\end{aligned}$$

Substituting the above and the FOCs into the citizen indifference condition gives

$$\begin{aligned}
&z = \Phi\left(\sqrt{\sigma_\varepsilon+\sigma_\theta}(\lambda\hat{x}+(1-\lambda)y-\hat{\theta})\right)c \\
&z_k - \frac{p}{2} = \Phi\left(\sqrt{\sigma_\varepsilon+\sigma_\theta}\left(\frac{\sigma_\varepsilon}{\sigma_\varepsilon+\sigma_\theta}\left[\frac{\Phi^{-1}(\hat{\theta})}{\sqrt{\sigma_\varepsilon}}+\hat{\theta}\right]+\frac{\sigma_\theta}{\sigma_\varepsilon+\sigma_\theta}\left[y_k+\frac{p}{2}\right]-\hat{\theta}\right)\right)c \\
&z_k - \frac{p}{2} = \Phi\left(\frac{\sqrt{\sigma_\varepsilon+\sigma_\theta}}{\sigma_\varepsilon+\sigma_\theta}\left(\sigma_\varepsilon\left[\frac{\Phi^{-1}(\hat{\theta})}{\sqrt{\sigma_\varepsilon}}+\hat{\theta}\right]+\sigma_\theta\left[y_k+\frac{p}{2}\right]-(\sigma_\varepsilon+\sigma_\theta)\hat{\theta}\right)\right)c \\
&z_k - \frac{p}{2} = \Phi\left(\frac{1}{\sqrt{\sigma_\varepsilon+\sigma_\theta}}\left(\frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon}}\left[\Phi^{-1}(\hat{\theta})+\sqrt{\sigma_\varepsilon}\hat{\theta}\right]+\sigma_\theta\left[y_k+\frac{p}{2}\right]-(\sigma_\varepsilon+\sigma_\theta)\hat{\theta}\right)\right)c \\
&z_k - \frac{p}{2} = \Phi\left(\frac{1}{\sqrt{\sigma_\varepsilon+\sigma_\theta}}\left(\sqrt{\sigma_\varepsilon}\left[\Phi^{-1}(\hat{\theta})+\sqrt{\sigma_\varepsilon}\hat{\theta}\right]+\frac{\sqrt{\sigma_\varepsilon}}{\sqrt{\sigma_\varepsilon}}\sigma_\theta\left[y_k+\frac{p}{2}\right]-\frac{\sqrt{\sigma_\varepsilon}(\sigma_\varepsilon+\sigma_\theta)}{\sqrt{\sigma_\varepsilon}}\hat{\theta}\right)\right)c \\
&z_k - \frac{p}{2} = \Phi\left(\frac{\sqrt{\sigma_\varepsilon}}{\sqrt{\sigma_\varepsilon+\sigma_\theta}}\left(\Phi^{-1}(\hat{\theta})+\sqrt{\sigma_\varepsilon}\hat{\theta}+\frac{\sigma_\theta\left[y_k+\frac{p}{2}\right]}{\sqrt{\sigma_\varepsilon}}-\frac{\sigma_\varepsilon+\sigma_\theta}{\sqrt{\sigma_\varepsilon}}\hat{\theta}\right)\right)c \\
&z_k - \frac{p}{2} = \Phi\left(\frac{\sqrt{\sigma_\varepsilon}}{\sqrt{\sigma_\varepsilon+\sigma_\theta}}\left(\Phi^{-1}(\hat{\theta})+\frac{\sigma_\varepsilon-\sigma_\varepsilon-\sigma_\theta}{\sqrt{\sigma_\varepsilon}}\hat{\theta}+\frac{\sigma_\theta\left[y_k+\frac{p}{2}\right]}{\sqrt{\sigma_\varepsilon}}\right)\right)c \\
&z_k - \frac{p}{2} - \Phi\left(\frac{\sqrt{\sigma_\varepsilon}}{\sqrt{\sigma_\varepsilon+\sigma_\theta}}\left(\Phi^{-1}(\hat{\theta})+\frac{\sigma_\theta}{\sqrt{\sigma_\varepsilon}}(y_k+\frac{p}{2}-\hat{\theta})\right)\right)c = 0 \\
&U(\hat{\theta},\sigma_\varepsilon,\sigma_\theta,y) = 0
\end{aligned}$$

Notice that the limits of the payoff difference in terms of θ are

$$\begin{aligned}
\lim_{\theta\to 0} U(\theta,\sigma_\varepsilon,\sigma_\theta,y) &= z_k - \frac{p}{2} - \Phi\left(\frac{\sqrt{\sigma_\varepsilon}}{\sqrt{\sigma_\varepsilon+\sigma_\theta}}\left(\underbrace{\Phi^{-1}(0)}_{-\infty} + \frac{\sigma_\theta}{\sqrt{\sigma_\varepsilon}}\left[y_k + \frac{p}{2}\right]\right)\right)c \\
&= z_k - \frac{p}{2} - \Phi(-\infty)c \\
&= z_k - \frac{p}{2} > 0 \\
\lim_{\theta\to 1} U(\theta,\sigma_\varepsilon,\sigma_\theta,y) &= E[b] - \Phi\left(\frac{\sqrt{\sigma_\varepsilon}}{\sqrt{\sigma_\varepsilon+\sigma_\theta}}\left(\underbrace{\Phi^{-1}(1)}_{\infty} + \frac{\sigma_\theta}{\sqrt{\sigma_\varepsilon}}(y_k + \frac{p}{2} - 1)\right)\right)c \\
&= z_k - \frac{p}{2} - \Phi(\infty)c \\
&= z_k - \frac{p}{2} - c < 0
\end{aligned}$$

Therefore, we can conclude that the function $U(\theta,\sigma_\varepsilon,\sigma_\theta,y)$ is continuous in θ on $[0,1]$, so that $U(0,\sigma_\varepsilon,\sigma_\theta,y)$ and $U(1,\sigma_\varepsilon,\sigma_\theta,y)$ have different signs. Equation $U(\theta,\sigma_\varepsilon,\sigma_\theta,y) = 0$ thus must have at least one solution, denoted by $\hat{\theta}$. Moreover, by assuming that

$$\sqrt{2\pi} > \frac{\sigma_\theta}{\sqrt{\sigma_\varepsilon}}$$

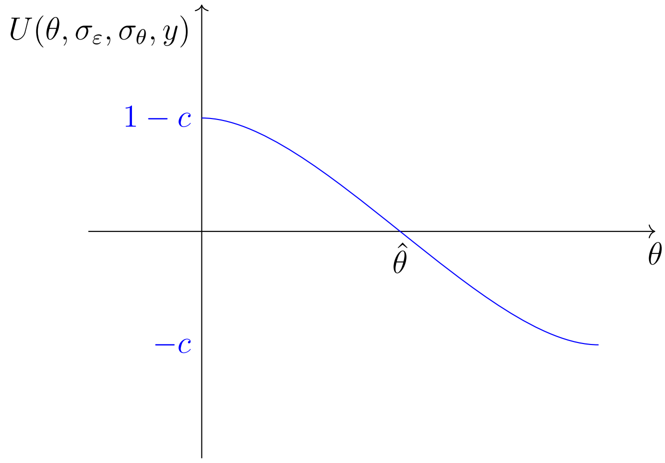
We can conclude that $U(\theta,\sigma_\varepsilon,\sigma_\theta,y)$ is strictly monotonic, and the solution is unique[11]

$$\begin{aligned}
\min_\theta \frac{1}{\phi(\Phi^{-1}(\theta))} &= \sqrt{2\pi} > \frac{\sigma_\theta}{\sqrt{\sigma_\varepsilon}} \Rightarrow \\
\frac{\partial U(\theta,\sigma_\varepsilon,\sigma_\theta,y)}{\partial\theta} &= -\phi\left(\frac{\sqrt{\sigma_\varepsilon}}{\sqrt{\sigma_\varepsilon+\sigma_\theta}}\left(\Phi^{-1}(\theta) + \frac{\sigma_\theta}{\sqrt{\sigma_\varepsilon}}(y-\theta)\right)\right)\left[\frac{\sqrt{\sigma_\varepsilon}}{\sigma_\varepsilon+\sigma_\theta}\left(\frac{1}{\phi(\Phi^{-1}(\theta))} - \frac{\sigma_\theta}{\sqrt{\sigma_\varepsilon}}\right)\right] \\
&< 0
\end{aligned}$$

[11] Below, I use inverse function theorem and the following derivation:

$$\begin{aligned}
\operatorname{argmin}_\theta \frac{1}{\phi(\Phi^{-1}(\theta))} &= \operatorname{argmax}_\theta \phi(\Phi^{-1}(\theta)) \Rightarrow \Phi^{-1}(\theta) = 0 \Rightarrow \theta = \Phi(0) = 0.5 \\
\min \frac{1}{\phi(\Phi^{-1}(\theta))} &= \frac{1}{\phi(\Phi^{-1}(0.5))} \approx 2.506628 = \sqrt{2\pi}
\end{aligned}$$

Figure E1: Payoff difference $U(\theta, \sigma_\varepsilon, \sigma_\theta, y)$



Comparative statics Neither $\hat{\theta}$ nor $\hat{x}$ have closed form solutions. But given that $\partial U(\theta, \sigma_\varepsilon, \sigma_\theta, y)/\partial\theta < 0$, we can use IFT to derive some comparative statics. To do this, we derive the following partial derivatives

$$\frac{\partial U(\theta, \sigma_\varepsilon, \sigma_\theta, y)}{\partial c} = -\Phi(\cdot) < 0$$
$$\frac{\partial U(\theta, \sigma_\varepsilon, \sigma_\theta, y)}{\partial p} = -\frac{1}{2} - \phi(\cdot)\frac{\sigma_\theta c}{2\sqrt{\sigma_\varepsilon + \sigma_\theta}} < 0$$
$$\frac{\partial U(\theta, \sigma_\varepsilon, \sigma_\theta, y)}{\partial y_k} = -\phi(\cdot)\frac{\sigma_\theta c}{\sqrt{\sigma_\varepsilon + \sigma_\theta}} < 0$$
$$\frac{\partial U(\theta, \sigma_\varepsilon, \sigma_\theta, y)}{\partial z_k} = 1 > 0$$

Now it is easy to see how c, p, y_k and z_k each affect the regime threshold

$$\frac{\partial\hat{\theta}}{\partial c} = -\frac{\frac{\partial U(\theta,\sigma_\varepsilon,\sigma_\theta,y)}{\partial c}}{\frac{\partial U(\theta,\sigma_\varepsilon,\sigma_\theta,y)}{\partial\theta}} < 0, \qquad \frac{\partial\hat{\theta}}{\partial p} = -\frac{\frac{\partial U(\theta,\sigma_\varepsilon,\sigma_\theta,y)}{\partial p}}{\frac{\partial U(\theta,\sigma_\varepsilon,\sigma_\theta,y)}{\partial\theta}} < 0$$
$$\frac{\partial\hat{\theta}}{\partial y_k} = -\frac{\frac{\partial U(\theta,\sigma_\varepsilon,\sigma_\theta,y)}{\partial y_k}}{\frac{\partial U(\theta,\sigma_\varepsilon,\sigma_\theta,y)}{\partial\theta}} < 0, \qquad \frac{\partial\hat{\theta}}{\partial z_k} = -\frac{\frac{\partial U(\theta,\sigma_\varepsilon,\sigma_\theta,y)}{\partial z_k}}{\frac{\partial U(\theta,\sigma_\varepsilon,\sigma_\theta,y)}{\partial\theta}} > 0$$

The interpretation is straightforward: an increase in either the participation cost, censorship penalty or the aspired signal of government strength lower the regime threshold, allowing weaker regimes to survive. Fiercer government oppression naturally lowers the possibility of a regime change. Also, repression of free media or a willing pro-regime bias makes the incumbent regime stronger. Finally, an increase in agitation decreases the space of surviving regimes, by persuading or pressuring more people to join the cause.

Because of the nice expression for $\hat{x}$, the results regarding regime threshold carry right through for the participation threshold, too

$$\frac{\partial \hat{x}}{\partial c} = \underbrace{\frac{1}{\sqrt{\sigma_\varepsilon}\phi\left(\Phi^{-1}(\hat{\theta})\right)}}_{>0} \frac{\partial \hat{\theta}}{\partial c} + \frac{\partial \hat{\theta}}{\partial c} < 0$$

$$\frac{\partial \hat{x}}{\partial p} < 0, \qquad \frac{\partial \hat{x}}{\partial y_k} < 0, \qquad \frac{\partial \hat{x}}{\partial z_k} > 0$$