The Social Costs of Digital vs. In-person Surveillance: Experimental and Observational Evidence from China

Xu Xu*

(Draft version. Please do not circulate.)

Abstract

Autocratic regimes employ surveillance for political control. Traditionally, states rely on human informants to penetrate society to collect information. In recent decades, autocracies increasingly adopt advanced information technologies for digital surveillance. This paper examines the social costs of digital surveillance versus in-person surveillance. I argue that both types of surveillance deter civic participation because citizens fear targeted repression. However, digital surveillance does not entail human-agent intrusion into private lives, and therefore is less likely to lower interpersonal trust and regime legitimacy. I find consistent evidence using a survey experiment with over 500 students in two universities in North and West China. The external validity of the experimental findings is established using the Chinese General Social Survey and an interrupted time-series design that exploits an exogenous shock to the intensity of digital surveillance and repression caused by the 2015 Tianjin Explosions in China.

*For helpful comments, I would like to thank Xun Cao, Volha Charnysh, Martin Dimitrov, Matt Golder, Luke Keele, Genia Kostka, Margaret Roberts, Joseph Wright, Yiqing Xu, Boliang Zhu; participants at the MPSA conference, the annual meeting of the Society for Political Methodology, and department seminars at Penn State. Finally, I would like to thank Xin Jin for her helpful comments and unwavering support. All remaining errors are my own.

"WIR SIND ÜBERALL (WE ARE EVERYWHERE)"

— Stasi, German Democratic Republic

1 Introduction

Autocratic regimes suppress civic engagement and political participation (Almond and Verba 1963). Most dictatorships use some form of surveillance for political control. Traditionally, autocratic regimes rely on human informants to spy on citizens. In the German Democratic Republic (East Germany), the Stasi (short for *Staatssicherheitsdienst*, or State Security Service) employed over 90 thousand employees and nearly 170 thousand informants who collected vast amounts of information used to intimidate the citizens and eliminate regime opposition. Counting part-time informants, the Stasi had one agent per 6.5 people (Gieseke 2014). More recently in Iraq, the Ba'th Party's surveillance apparatus managed to penetrate many parts of Iraq society. For example, the regime punished citizens for failing to turn in an associate who committed even minor transgressions (Faust 2015). In societies penetrated by human agents who gather information directly from observing and interacting with their targets, citizens are atomized from one another. Citizens fear revealing their true anti-regime sentiments, and low levels of interpersonal trust are common (Blaydes 2018). The lack of common knowledge about support for the government and low-levels of trust hinder citizens' ability to coordinate uprisings against the regime (Chwe 2013). While human surveillance may help prevent anti-regime coordination and protest, it also undermines the regime's legitimacy and the prospects for economic development (Fukuyama 1995).

In recent decades, autocracies increasingly turn to *digital* surveillance for population control. For example, the Iranian and Syrian governments have developed sophisticated digital surveillance systems to track and identify opposition members; other authoritarian countries such as Saudi Arabia, Vietnam, and Russia are learning from China to build digital technologies for domestic surveillance and censorship (Gunitsky 2015). Unlike traditional surveillance, digital surveillance does not rely on human agents to collect information di-

rectly fro. Instead, citizens transmit information in an electronic form on the internet that can be accessed and analyzed by automatic algorithms and machines with minimal human assistance. Given the rapid rise in digital surveillance around the world, it is important to understand how this type of surveillance differs from traditional, informant-based surveillance in shaping civic engagement, political participation, and interpersonal trust in authoritarian countries.

I argue that both types of surveillance deter free speech and civic participation. Surveillance helps dictators selectively repress citizens who are most likely pursue anti-regime activities. When faced with the prospect of repression, citizens under surveillance have an incentive to falsify their preferences and refrain from expression (Kuran 1991). Surveillance also deters protest coordination because under repression threats, individuals expect higher cost of participation – for themselves and for other – and anticipate fewer people to participate. A key feature of the Stasi's surveillance strategy were "silent" methods of repression rather than overt persecution by the police (Knabe 1999). Despite these similarities, digital surveillance and in-person surveillance have different implications for interpersonal trust and regime legitimacy. For traditional, in-person surveillance to work, surveillance agents must penetrate citizens' close social networks to collect information and identify dissidents. Such close human contact pits citizens against each other, diminishing interpersonal trust and regime support. On the contrary, digital surveillance does not entail human-agent intrusion into citizens' private lives, and is less likely to decrease trust and regime legitimacy.

I use a survey experiment with a sample of over 500 students in two universities in North and West China to examine the social costs of digital surveillance and in-person surveillance. Respondents are randomly assigned to read information about a digital surveillance scenario, an in-person surveillance scenario, and a control scenario. I find that both types of surveillance negatively influence political expression and intent to protest. In addition, both types of surveillance reduce respondents' beliefs about how many other individuals would participate in "anti-regime" collective action. Moreover, causal mediation analysis suggests that

both types of surveillance discourage protest participation by shaping people's beliefs about other individuals' participation in protest rather by reducing interpersonal trust. These findings suggest that both human and digital surveillance aid authoritarian survival by deterring coordinated anti-regime action. However, in-person surveillance significantly reduces interpersonal trust and regime legitimacy whereas digital surveillance has no effect on legitimacy and a small, negative effect on trust. This suggests that digital surveillance does not yield the same potential costs to the regime – in the form of decreased interpersonal trust and regime legitimacy – that human surveillance entails. Thus, from the perspective of regime control, digital surveillance yields many of the same benefits as human surveillance (decreasing citizen coordination) without some of the costly byproducts of surveillance, namely decreased interpersonal trust and regime legitimacy.

To establish the external validity, I use the 2015 Chinese General Social Survey on a nationally representative sample of nearly 11 thousand respondents and an interrupted time series design that exploits an exogenous shock to digital surveillance caused by the 2015 Tianjin Explosions in China. At midnight 12 August 2015, a series of massive explosions in the port of Tianjin killed 173 people and injured nearly a thousand (Merchant 2017). Online surveillance and censorship increased tenfold in China immediately afterwards (Dou 2015). By comparing respondents surveyed just before the explosions with those surveyed just afterwards, I show that digital surveillance decreases citizens' faith in free speech and petitioning the government but that increased digital surveillance has no effect on interpersonal trust. To further confirm that digital surveillance is the mechanism underpinning these results, I test provincial-level surveillance intensity, using the number of pilot counties for China's digital surveillance system (the Golden Shield Project, developed by the Ministry of Public Security). I show that citizens feel less secure about expressing free speech and petitioning the government in provinces with more intensive surveillance and censorship.

Many early proponents of Internet development believed that information technologies would spread freedom and spur democratization. Yet, two decades after the advent of the

digital era, we have not observed widespread authoritarian collapse. A growing body of literature on ICT and authoritarian survival explores how authoritarian governments use the Internet and ICT to censor and repress online expressions (King, Pan and Roberts 2013), collect information about citizen preferences (Gunitsky 2015), monitor local politicians (Qin, Strömberg and Wu 2017), distract or guide public opinion (King, Pan and Roberts 2017; Roberts 2018), and identify demonstrators and political opponents (Gunitsky 2015). This paper contributes to the literature by emphasizing the role of digital surveillance in deterring mass mobilization. Indeed, truthful communication and beliefs in others' participation may be necessary for citizens to coordinate a successful protest (Chwe 2013; Edmond 2013). For this reason, this paper explores how digital surveillance prevents mobilization in authoritarian countries by discouraging expression, reducing willingness to protest, and decreasing beliefs about others' willingness to participate in anti-regime collective action.

This paper also examines how surveillance influences interpersonal trust and civic participation – two cornerstones of free societies and the building blocks of open access markets. Interpersonal trust is essential for facilitating trade and well-functioning market economies (Greif 1989; Algan and Cahuc 2010). Civic participation, particularly as the basis for coordinating citizens' anti-government behavior, is a central part of toppling dictatorships peacefully and establishing free and democratic societies (Kendall-Taylor and Frantz 2014). Though studies of both topics are numerous (e.g., Almond and Verba 1963; Putnam, Leonardi and Nanetti 1994), few have considered surveillance, especially digital surveillance, as a determinant of trust and civic participation. In addition, recent studies find that autocratic rule decreases individuals' trust and civic engagement even after citizens migrate to democratic countries for generations (Xu and Jin 2018), but the exact sources of social distrust and isolation in autocracies are not well understood. This paper contributes to this literature by identifying how autocratic rule damages the fabric of society: government digital and human surveillance.

Finally, this paper informs historical research on traditional, informant-based surveil-

lance, especially the Stasi in East Germany (e.g., Bruce 2010; Gieseke 2014; Blaydes 2018). While the pernicious social consequences of human surveillance are well-documented, quantitative research on state surveillance is surprisingly thin. In addition, a core inferential challenge in previous work is the endogeneity that state surveillance is more likely to occur in low-trust societies. This paper provides systematic causal evidence linking government surveillance to individual social behavior.[1] Moreover, this research compares the social costs of digital surveillance to the costs of human surveillance. My findings therefore have important implications for understanding authoritarian control in the digital age. In particular, digital surveillance, like human surveillance, deters political expression and protest participation, but this digital repression, unlike human informants, does so without lowering interpersonal trust and regime legitimacy. The reduced social costs of surveillance in the digital age imply that dictators might be more likely to rely on digital surveillance to control society.

2 Surveillance, Trust, and Civic Participation

Surveillance is a common tool for information collection and political control in dictatorships (Greitens 2016). Unlike democratic leaders, dictators are inherently uninformed because citizens in dictatorships often hide their true anti-regime sentiments when faced with the prospect of state repression (Kuran 1991). To gauge public opinion for policymaking and to contain threats before they spread, dictators historically rely on human security agents and/or informants to collect information from citizens. In the digital age, information technologies expand dictators' information-collection toolkit. As citizens move to the electronic media for socializing, networking, communicating, shopping, and expressing opinions, computers and algorithms replace human agents as tools for governments to collect information from citizens. This shift from human to digital surveillance has important implications for interpersonal trust and civic participation.

[1] See Lichter, Loeffler and Siegloch (2019) for a recent example.

2.1 In-person Surveillance

Human societies have a long tradition of in-person surveillance. In BC 839, King Li of Zhou Empire in China asked his wizards to spy on the people and kill those who criticized his tyranny (Zuo 1998). The Jinyiwei (Embroidered Uniform Guard) of the Ming Dynasty was founded in the 1360s by the Hongwu Emperor and served as the dynasty's secret police until the collapse of Ming rule in 1644. In Europe, secret police organizations emerged after the French Revolution in 18th-century; and Hitler's regime in Germany (1933–1945) utilized the Gestapo to eliminate opponents. East Germany (1945–1990) created the Stasi, with unparalleled social penetration. Other dictatorships such as Iraq under Saddam Hussein, Chile under Pinochet, Peru under Fujimori, Philippines under Marcos, and North Korea also used large secret police organizations to control society (Greitens 2016; McMillan and Zoido 2004). In-person surveillance in dictatorships shapes individuals' social behavior via betrayal and sabotage.

Societal Penetration and Betrayal. To obtain precise information about opposition groups, a traditional surveillance apparatus requires a large body of security agents and informants to penetrate citizens' social networks and private lives. For example, the Stasi, at its peak, employed nearly 90 thousand employees and over 170 thousand informants, comprising roughly 1.6 percent of the population in East Germany (Gieseke 2014). Other regimes such as the Kim regime in North Korea and the Republic of China (under Chiang Kai-shek and Chiang Ching-kuo) surveilled their citizens with massive secret police organizations (Greitens 2016). Informants are ordinary citizens but secretly gather information within their professional and social networks by betraying the trust of friends, neighbors, colleagues, relatives, and even family members (Ash 1998). Societal penetration and betrayal thus generate a widespread atmosphere of suspicion and distrust among citizens (Blaydes 2018). Linking surveillance operations to the government, citizens may blame the regime for their suffering under surveillance, thereby decreasing regime legitimacy.

Self-interested Human Agents. In-person surveillance relies on self-interested hu-

man agents and informants who use subjective assessment of other citizens' loyalty to the regime and may maliciously target "innocent" people to resolve personal disputes (Kalyvas 2006). To provide information, informants may demand benefits from the regime, such as government jobs, opportunities to travel abroad, or monetary compensation. In addition, potential informants may be "tricked" to provide service to clear up their own "blemishes" (Ash 1998). Due to the temptation of rewards or the urge to prove their innocence, self-interested agents may misreport or sabotage their fellow citizens. The potential power abuse by agents and informants further foment distrust and anti-regime sentiments in society.

These mechanisms suggest that in-person surveillance reduces interpersonal trust and regime legitimacy.

2.2 Digital Surveillance

Since the advent of the information era, dictatorships increasingly adopt digital surveillance for social control. By 2015, more than twelve authoritarian countries have or are developing digital tools for domestic surveillance (Valentino-DeVries, Vo and Yadron 2015). Relatively cheap spyware available on international markets has made digital surveillance tools accessible to the poorest dictatorships. Authoritarian governments use malware to spy on opposition leaders and journalists (Deibert 2017), collect metadata from social media to keep tabs on political opponents (Qin, Strömberg and Wu 2017), and employ high-resolution digital cameras and facial recognition technologies to identify dissidents (Liu and Wang 2017). Recent advances in artificial intelligence detect suspicious movements in crowds, identify thousands of people at once, and recognize citizens who attempt to conceal their identities by wearing hats, sunglasses, or scarves to cover their faces (Intel 2017; Singh et al. 2017).

Unlike in-person surveillance, digital surveillance does not entail human intrusion into citizens' private lives and thus may be less recognizable to citizens and less likely to breed back. In the digital age, citizens communicate and spend substantial time online, leaving

personal digital information for governments to access and analyze. Digital surveillance relies on Internet infrastructures such as computers, software, algorithms, cameras, cables, routers, servers, and data storage centers, rather than human informants. Further, digital surveillance allows governments to monitor a large population and reach the most private part of people's lives with minimal human assistance,[2] while computers and algorithms yield more accurate and objective data than human agents who often misreport or intentionally sabotage "loyal" citizens for personal gain. Thus, the absence of human betrayal and sabotage in the operation of digital surveillance suggests that it is less likely to reduce interpersonal trust and regime legitimacy than in-person surveillance.[3] This yields the following testable implications.

Trust hypothesis: Digital surveillance is less likely to reduce interpersonal trust than in-person surveillance.

Legitimacy hypothesis: Digital surveillance is less likely to reduce regime legitimacy than in-person surveillance.

2.3 Surveillance, Repression, and Civic Participation

Surveillance, whether digital or in-person, discourages civic participation because it entails preemptive, targeted repression against its target (Dimitrov and Sassoon 2014). In dictatorships where meaningful elections and other representative channels of political expression are often unavailable, civic participation may entail petitions, protests, and even violent revolts. These actions disrupt social order and may threat autocratic survival. Dictators thus use repression as a strategy to ensure political stability and avoid revolution (Wintrobe 2000). While dictators may employ "high-intensity", overt persecution against opponents,

[2]For example, digital cameras allow governments to analyze population who do not use the Internet; Data from search engines reveal people's private preferences such as pornography that they would not even tell their most close friends.

[3]Existing research on surveillance and public opinion also finds that digital surveillance does not reduce regime legitimacy because citizens are willing to sacrifice privacy for security.

more frequently they rely on "low-intensity", preemptive repression such as harassment, intimidation, or psychological coercion to weaken opposition challenges before they become strong (Sullivan 2016; Ritter and Conrad 2016). However, effective preemptive repression relies on accurate information about citizens' anti-regime preferences to identify opponents. Surveillance enables dictators to find dissidents for targeted repression, thereby discouraging citizens' anti-regime political expression and protest participation.[4] This suggests that, empirically, both types of surveillance should decrease observed political expression and protest participation.

Expression hypothesis: Both in-person and digital surveillance deter political expression.

Protest hypothesis: Both in-person and digital surveillance deter protest participation.

To mount a protest, participants must coordinate their actions (Chwe 2013). Surveillance may deter protest coordination via two channels. First, interpersonal trust may help induce protest participation by increasing individuals' belief that protest participation will be *safe* and *worthwhile* (Benson and Rochon 2004). Surveillance lowers interpersonal trust, thereby reducing individuals' willingness to participate. Second, strategic considerations are another important determinant of protest participation: an individual's behavior is shaped by beliefs about the participation of others (Edmond 2013). Thus, surveillance may also deter participation by influencing individuals' beliefs about how many others will participate. The following hypotheses examines these two potential channels through which surveillance deters protest participation.

Trust-protest hypothesis: Surveillance deters protest participation by lowering inter-personal trust among citizens.

Coordination-protest hypothesis: Surveillance deters protest participation by influencing

[4]Dictators may strategically operate large-scale state surveillance systems to silence expression and deter mobilization. The key feature of the Stasi's surveillance strategy was the use of "silent" methods of repression rather than legal persecution by the police (Knabe, 1999).

individuals' belief about others' participation.

3 Experimental Design

To test the hypotheses, I pursue two strategies. First, I conducted an in-the-field survey experiment with 539 students in March 2019 to compare the social consequences of in-person surveillance with those of digital surveillance. Second, I address the external validity of the experimental findings by analyzing a nationally representative survey in an interrupted time-series (ITS) setting. This section explains the details of the experimental design. The ITS design comprises Section 5.

3.1 Design and Randomization

I recruited students in two universities in North and West China to broaden sample representativeness. As Figure 1 shows, the home provinces of the student sample cover most regions in China. I choose university students because they are the most active social group in political participation. For example, in the 1960s, students initiated the anti-war movement and actively participated in Vietnam war protests in the U.S. (Moore 1999). In 1989, hundreds of thousands student protesters occupied the Tiananmen square to demand democracy in China, sparking large-scale student protests throughout the country (Zhao 2004). During the 2014 Hong Kong protests, students were also at the heart of the "Umbrella Revolution" protests (Cantoni et al. 2019). If we can find a deterrence effect of surveillance on civic participation in the demographic group most likely to protest, we are likely to observe a stronger effect in less active demographics. Thus, examining a student sample sheds light on a much larger population.

I use a within-subject design to compare interpersonal trust, civic participation, and regime support among three groups: a treated group with in-person surveillance, a treated group with digital surveillance, and a control group without surveillance.[5] To assess possible

[5]I conducted a pilot study using a between-subject design on a sample of 214 students in one university in East China. Statistical analysis found large variation in the trust and

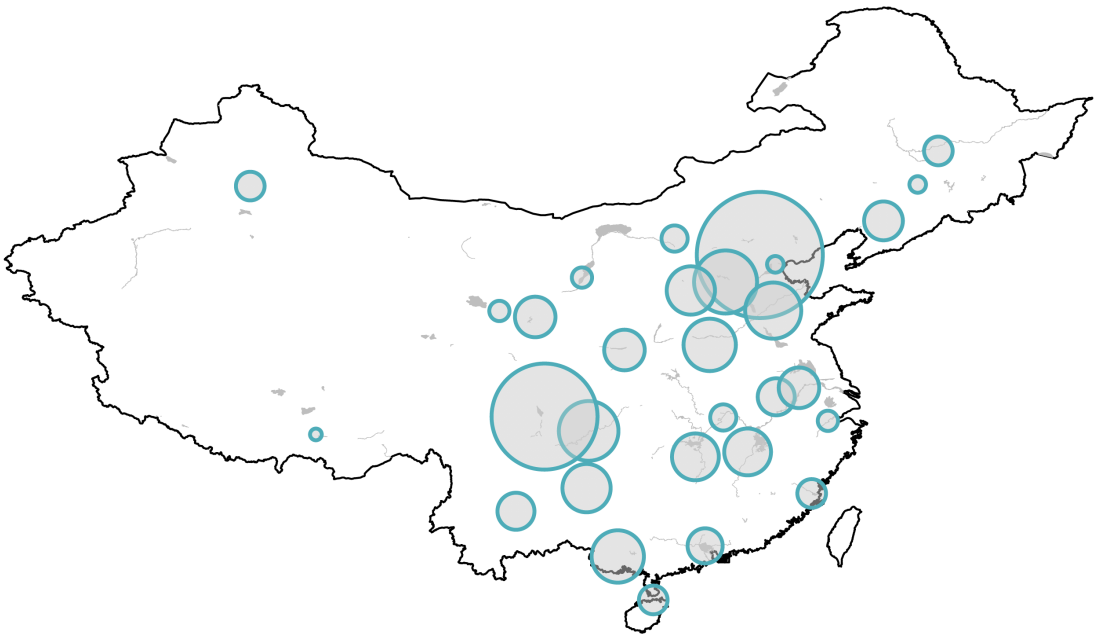


Figure 1: Distribution of Respondents by Home Province

experimenter demand effects of the within-subject design, I implement a between-subject design for comparison. This joint design scheme entails two design groups with a total of six experimental groups. Each student has a 2/3 chance of assignment to the within-subject design group and a 1/3 chance of assignment to the between-subject design group. Within each design group, students have an equal chance of assignment to the control condition or one of the two treatment conditions. In total, 353 students were randomly assigned to three groups in the within-subject design, and 186 students were randomly assigned to three groups in the between-subject design.

During the survey, all respondents first answer background questions about age, gender, income, party affiliation, social distrust, civil participation, and media usage. They

participation variables, implying a very large sample would be necessary to estimate statistically significant differences. As we know, trust and civic participation are intrinsic values (Putnam, Leonardi and Nanetti 1994). With a relatively small sample size, I implement a within-subject design that differentiates respondents' intrinsic values, following Wiswall and Zafar (2014).

then read a descriptive vignette conveying information about protest participation that is related to their campus life. Respondents in the within-subject design group first answer questions about willingness to make public political expressions and protest, trust toward fellow students, as well as support for university authority. After receiving information about the school authority's in-person surveillance, digital surveillance, or no surveillance operation, the respondents answer the same questions again. These pre- and post-tests allow me to differentiate out respondents' intrinsic attitudes. Alternatively, respondents in the between-subject design group directly receive information about surveillance without the pretest questions. Table 1 shows the structure of the design.

Table 1: Experimental Design for In-person vs. Digital Surveillance

	Within-Subjects Design			Between-Subjects Design		
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Assignment:	Control	Treatment 1	Treatment 2	Control	Treatment 1	Treatment 2
Background Qs:	Yes	Yes	Yes	Yes	Yes	Yes
Scenario:	Yes	Yes	Yes	Yes	Yes	Yes
Pre-test:	Yes	Yes	Yes	–	–	–
Treatment:	No Surveil.	In-person	Digital	No Surveil.	In-person	Digital
Post-test:	Yes	Yes	Yes	Yes	Yes	Yes
N:	124	103	126	61	64	61

Table 2 reports the covariate balance across control and treatment groups on a number of background questions, including age, gender, family income, income satisfaction, party affiliation, membership in official university organizations, membership in student societies, community service, interest in discussing politics, media usage, social distrust, online expression. As shown in Table 2, randomization is successful and the treatment is balanced across all these covariates.

3.2 Treatments and Measures

Protest Scenario and Surveillance Treatments

Table 3 shows the vignette for the hypothetical scenario in which students confront the university authority for an unfair policy involving changing dormitories. This scenario

Table 2: Balance Check, Within-Subject Sample

	obs.	Control	Human	Digital	p-value
Age	337	20.41	20.47	20.60	0.772
Female (F=1)	341	0.52	0.48	0.51	0.831
Income (1-9)	338	6.66	6.59	6.81	0.644
Income Sat. (0-10)	342	6.62	6.96	6.80	0.582
Party (Yes=1)	343	0.09	0.11	0.11	0.845
Offical Org. (Yes=1)	343	0.46	0.48	0.54	0.501
Stud. Org. (Yes=1)	342	0.64	0.67	0.61	0.700
Commu. Serv. (1-5)	343	2.52	2.61	2.60	0.710
Speech (1-5)	342	3.11	3.06	3.07	0.896
Media: News (1-5)	334	2.08	2.10	2.12	0.925
Media: TV (1-5)	337	3.02	2.91	2.91	0.638
Media: Phone (1-5)	343	4.72	4.73	4.75	0.896
Distrust (0-10)	342	4.22	4.57	3.81	0.175
Diss. Politics (1-5)	341	2.30	2.13	2.25	0.442

corresponds to real-world confrontations between citizens and the government. The setting in which students appeal to the Ministry of Education also mimics citizens' petitions to upper-level administrations for justice – a common phenomenon in authoritarian countries (Lorentzen 2013). To simulate real-world protest coordination, I remind respondents that more participants lead to a higher chance of petition success. In addition, I made punishment salient to remind the respondents that protest participation is costly. This punishment also helps respondents relate the surveillance treatments to repression. In short, this scenario is relevant to campus life to elicit students' truthful responses, while representing a common situation of contentious participation in authoritarian countries.

Table 4 lists the scenarios of in-person surveillance, digital surveillance, and no surveillance. The in-person surveillance scenario mimics the traditional, Stasi-style surveillance that employs human informants to spy on fellow citizens. The digital surveillance scenario is similar to real-world online surveillance conducted by authoritarian governments (Gunitsky 2015). Because respondents' prior experience of surveillance may influence their responses even if they do not receive any new information about surveillance, I specify in the control condition that "the university does not know who who participates in the protest" to anchor respondents' beliefs.

Measurement

Table 3: Scenario of Civic Participation

Imagining the following scenario:

Without holding any public hearings or conducting any opinion surveys among the students, the university unilaterally notifies you and other students who live in a new residence hall to move to one of the university's oldest and poorly maintained dormitories. Furthermore, your housing rates are not reduced. The new residence hall that you are currently living in will be freed up to accommodate an increased number of freshmen due to the university's recent enrollment expansion. You and other students in your residence hall are very upset with the university's decision and are complaining about this change. You and others are considering filing official complaints to the Ministry of Education who oversees your university, using the Ministry's online mailbox. The more students participate, the more likely you will push the ministry to change the university's decision. But once the university find out who participate, it may interrogate or even demerit the participants.

Table 4: Treatment Scenarios

Treatment 1: In-person Surveillance

The Ministry of Education's online mailbox can be filed anonymously. However, the university authority approached some of the students in your residence hall and promised them some benefits (you do not know whom those students are and what benefits will they receive). In exchange, those students agreed to secretly investigate and report the names of the students who participate in the filing process as well as the students who promoted this protest.

Treatment 2: Digital Surveillance

The Ministry of Education's online mailbox can be filed anonymously. However, the university authority can monitor students' online activities (on social media/apps, websites, forums, etc.) through the university's Internet servers to identify the students who participate in the filing process as well as the students who promoted this protest.

Control: No Surveillance

Because the Ministry of Education's online mailbox can be filed anonymously, the university does not know who participates in the protest.

In the survey, I ask respondents' willingness to express their discontent both in front of their fellow students and online. For protest participation, I elicit respondents' willingness to appeal to the Ministry of Education, as well as their beliefs regarding other students' participation (percentage points). The responses for willingness to express and protest are recorded on an ordinal scale from 1 to 4, with 4 indicating the most affirmative answer.

Interpersonal trust is measured by respondents' trust toward other students in the same residential hall on an ordinal scale from 0 to 10, with 10 indicating the highest level of trust.[6] Regime legitimacy is measured by the extent to which respondents approve most of the university's policies, regulations, and decisions, also on a scale from 0 to 10. Online Appendix A.1 documents the survey instruments in detail.

3.3 Implementation and Ethics

Implementation

Conducting the survey experiment on a potentially sensitive topic in the field circumvents state censorship that may be present in China-based online survey platforms. It also helps create trust and cooperation from respondents. More importantly, since one of the treatment conditions is online surveillance, it may induce respondents' self-censorship in online surveys. An in-the-field survey experiment avoids this problem because respondents answer questions on paper questionnaires.

The enumerators conducted the survey in dining halls and main roads between classroom buildings and residential halls. For a convenience sample, respondents were recruited in those areas to represent the student population better than in dormitories or classrooms because all students come to dining halls and main roads regardless of their major, gender, and year at university. Survey questionnaires require five to ten minutes to complete. Respondents were requested to complete the questionnaire independently to minimize potential spillover effects of the treatments. Each student received five Chinese Yuan (about 0.75 USD) as compensation. The six different versions of questionnaires were placed in a random order. The enumerators asked students whether they were willing to participate in an *anonymous* survey first, and if they agreed, the enumerators then presented the five-Yuan compensation

[6]Scholars find that the 11-point scales consistently outperformed the dichotomous counterpart for measuring trust in surveys (Krosnick and Fabrigar 1997; Lundmark, Gilljam and Dahlberg 2015).

to them and gave them the questionnaires in the random order. Roughly 50% of the students approached by enumerators agreed to participate. This response rate is within the normal range for a field survey. In addition, most of non-respondents refused before the enumerators explained the survey topic to them – their unwillingness to participate was thus not due to the content of the survey.[7] Thus, it is unlikely that non-response is related to potential outcomes that would bias results.

Ethical Considerations

As described in the acknowledgement, I sought and obtained research approval for the study from the Institutional Review Board (IRB) at Penn State. Given that this is essentially a five-minute opinion survey with minimal risk, it received an IRB exemption. I further take the following efforts to protect the rights and wellbeing of research participants and field staff.

First, I design the entire questionnaire to focus on students' campus lives without mentioning any sensitive political issues in China. This framing strategy reduces the risk of participating in the survey. In addition, I ask students' attitudes toward the school authority rather than the government. The hypothetical confrontation is between students and the school authority instead of between citizens and the state. Such questions are safe in the political context of China. During the recruitment and survey procedures, no student refused participation because of the content of the questionnaire. Second, I use an online petition instead of a street protest because stating participation in the former is less sensitive and safer for respondents, field staff, and the researcher. Third, in each university, I consulted with several faculty members and students to proofread the questionnaire to ensure it did not incur any risk to respondents. Finally, the survey is anonymous and the enumerators were requested to stay away when respondents answered questions. A questionnaire that does *not* collect personal information reduces the risk of a loss of confidentiality and any other potential risks to the subjects. These strategies not only reduce potential risks to

[7]The reasons for nonresponses include "no time", "hungry", "too busy", etc.

respondents, but also minimize respondents' social desirability bias and self-censorship in answering questions.

4 Experimental Findings

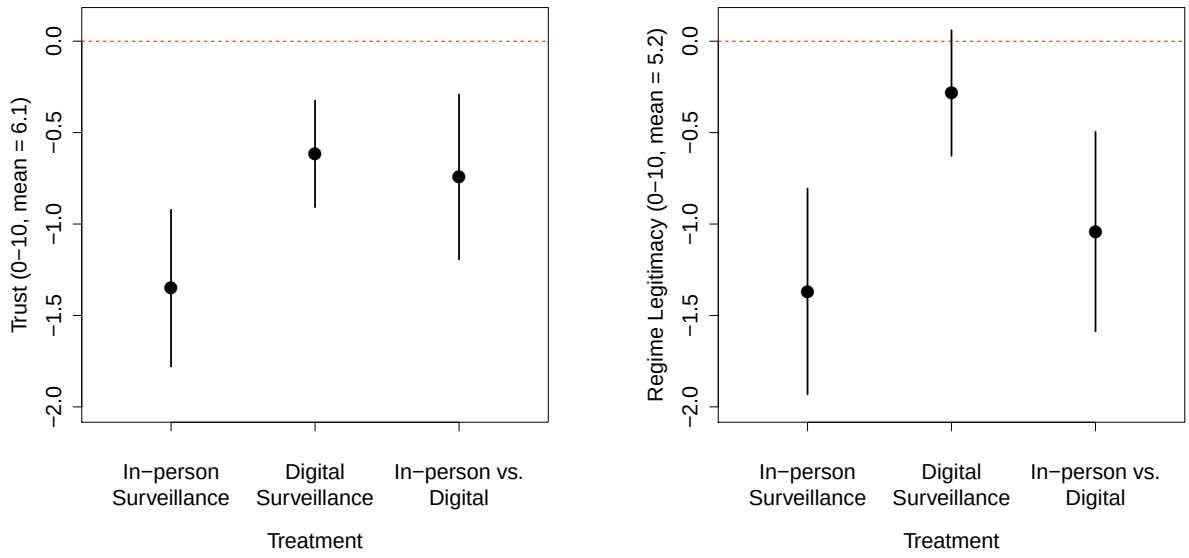
This section presents the results of the survey experiment based on the within-subject design. I take the differences between post-test answers and pre-test answers to generate the outcome variables of trust, regime legitimacy, expression, and protest participation. For each outcome variable, I first examine the effects of in-person surveillance and digital surveillance using no-surveillance as the comparison group. I then compare in-person surveillance with digital surveillance using the latter as the comparison group. The statistical inferences are based on standard comparisons of means using OLS estimation.[8]

4.1 Trust and Regime Legitimacy

I begin by presenting the effects of in-person and digital surveillance on interpersonal trust and regime legitimacy. My theory suggests that in-person surveillance reduces trust and legitimacy because it relies on self-interested human agents to penetrate into individuals' social networks and betray them to collect information. In contrast, digital surveillance does not entail human-agent intrusion of private lives, and is less likely to lower trust and regime legitimacy. In addition, compared with in-person surveillance, digital surveillance is less likely to reduce trust and legitimacy. The evidence from the survey experiment is consistent with these predictions.

As Figure 2.(a) shows, in-person surveillance largely reduces interpersonal trust. Recall that interpersonal trust is measured on a scale of 0 to 10, with 10 indicating highest level of trust. Given that the sample mean is 6.1 and the standard deviation is 2.1, a 1.4 decrease in trust scale is quite substantial (23%). Digital surveillance only slightly reduces trust (0.6 or

[8]As Table 2 shows that social distrust is not very balanced between three groups, I control for this variable in all specifications.



(a) Interpersonal Trust (b) Regime Legitimacy

(Notes: OLS estimates with 95% Confidence Intervals. See Online Appendix B.1 for the regression results underlying these figures.)

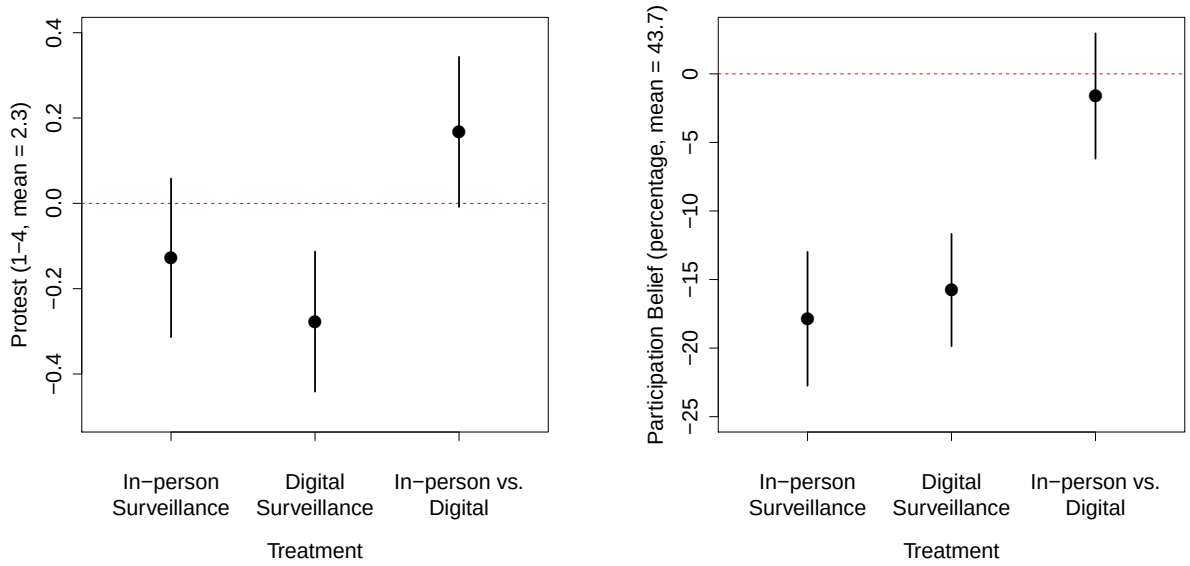
Figure 2: Trust and Legitimacy

10%). Both effects are statistically significant at 0.001 level. More importantly, compared with digital surveillance, in-person surveillance further reduces trust by 13%, and the effect is statistically significant. Figure 2.(b) examines regime legitimacy (on a scale of 0 to 10) and shows that in-person surveillance reduces legitimacy by a large margin (1.4 or 26%, significant at 0.001 level) whereas digital surveillance has a statistically insignificant effect. The negative effect of in-person surveillance is 19% larger than that of digital surveillance and is statistically significant at 0.001 level.

4.2 Civic Participation

I next examine how in-person and digital surveillance influence various aspects of civic participation. As discussed in the theoretical part, I expect both types of surveillance to discourage expression and protest participation because surveillance is associated with repression. Figure 3.(a) shows that both digital surveillance and in-person surveillance deters protest participation and the effect of digital surveillance is stronger. In particular, digital surveillance decreases respondents' willingness to participate by 0.28. Given that the

sample mean is 2.3 and the standard deviation is 1, the negative effect is substantial. In-person surveillance also negatively affects protest participation but the effect is statistically insignificant. This is likely due to my framing of the protest as an online petition instead of a street protest. In theory, digital surveillance deters online petition whereas in-person surveillance may not because citizens can hide their online activities from their friends, colleagues, or family members. Nevertheless, the negative effects of both types of surveillance are consistent with the theoretical predictions.



(a) Protest Participation (b) Blief in Others' Participation

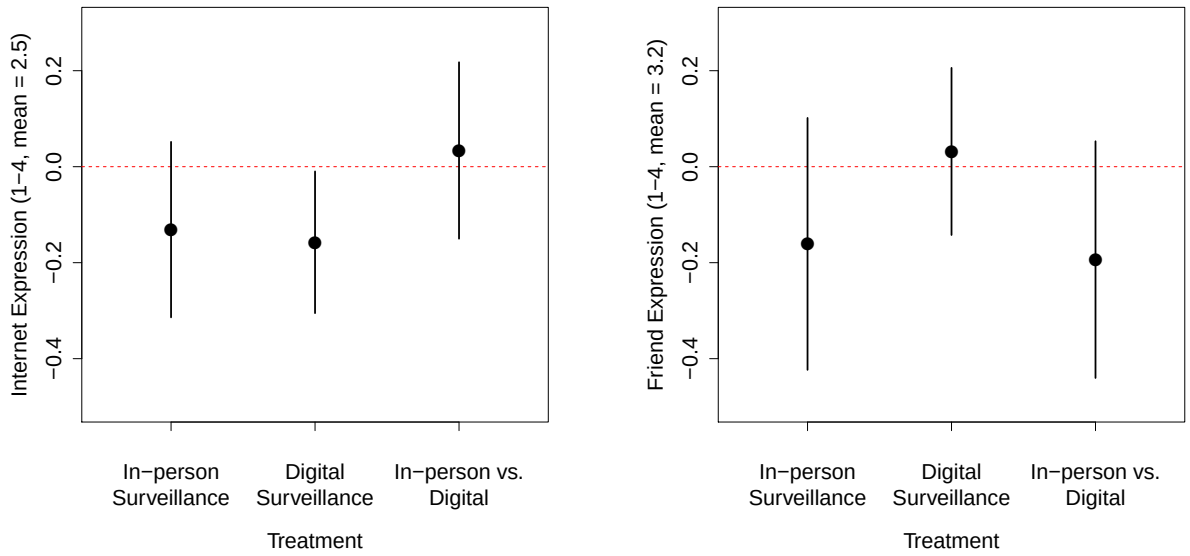
(Notes: OLS estimates with 95% Confidence Intervals. See Online Appendix B.1 for the regression results underlying these figures.)

Figure 3: Protest Participation and Beliefs

Figure 3.(b) presents the effects of surveillance on individuals' belief about others' protest participation. On average, respondents believe that 43.7% students in the residence hall will participate in the protest. As we can see, in-person and digital surveillance reduce the belief by 17.9 and 15.8 percentage points respectively, which are about 41% and 36% decreases. These large, negative effects are highly significant, and there is no statistically significant difference between two types of surveillance in affecting respondents' beliefs.

Figure 4 plots the effects of digital and in-person surveillance on online and offline expres-

sion. Both types of surveillance reduce respondents' willingness to express their discontent online and the effect of digital surveillance is statistically significant. There seems no discernible difference between two types of surveillance in deterring online expression. With regard to expression in front of fellow students, in-person surveillance has a negative but insignificant effect whereas the effect of digital surveillance is close to zero. The findings that digital surveillance discourages online expression but not offline expression meet our expectations.



(a) Online Expression (b) Expression to Fellow Students

(Notes: OLS estimates with 95% Confidence Intervals. See Online Appendix B.1 for the regression results underlying these figures.)

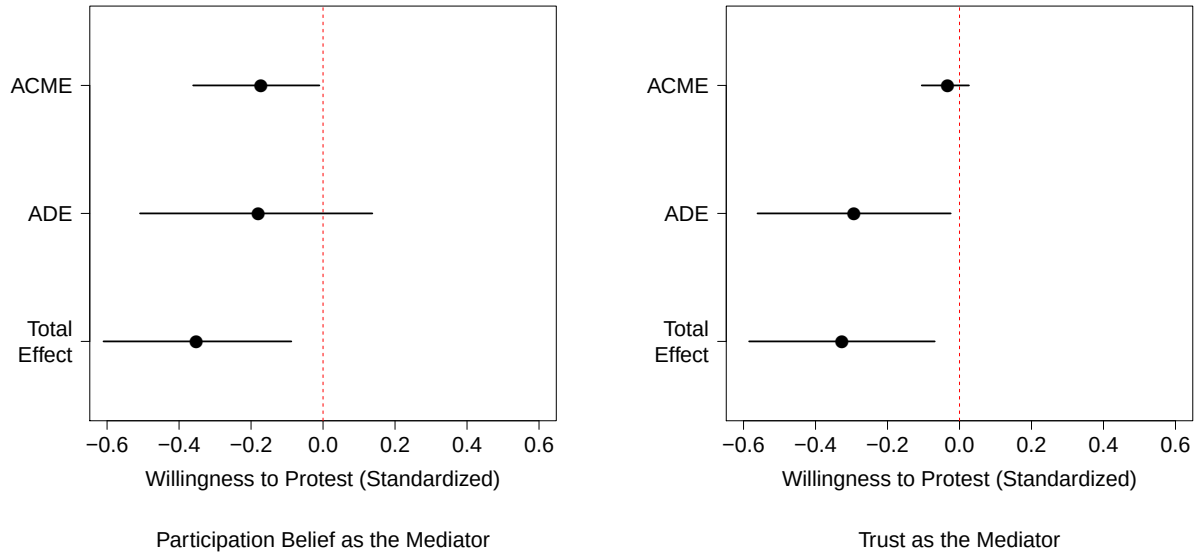
Figure 4: Online and Offline Expression

4.3 Causal Mediation Tests

I further examine the mechanisms through which surveillance deters protest participation using an approach of causal mediation analysis developed by Imai, Keele and Tingley (2010). As I mentioned in the theory section, both interpersonal trust and strategic consideration of others' participation could reduce individuals' willingness to protest. Thus, I use interpersonal trust and respondents' belief about others' turnout as mediators.

There are two major assumptions underlying the causal mediation test in identifying the mediation effect. First, there should be no unmeasured confounders between surveillance and willingness to protest. Second, there should be no unmeasured confounders between the mediator and willingness to protest. The first assumption holds since surveillance treatments are randomized. To satisfy the second assumption, I control for as many covariates as possible, including age, gender, family income, income satisfaction, party affiliation, membership in official school organization, membership in student organization, interest in discussing politics, media usage, and social distrust.

Figure 5 plots the results from the causal mediation analysis of digital surveillance on willingness to protest. Individuals' belief about others' participation has a positive and statistically significant average causal mediation effect (ACME). In contrast, the ACME of trust is close to 0 and statistically insignificant.

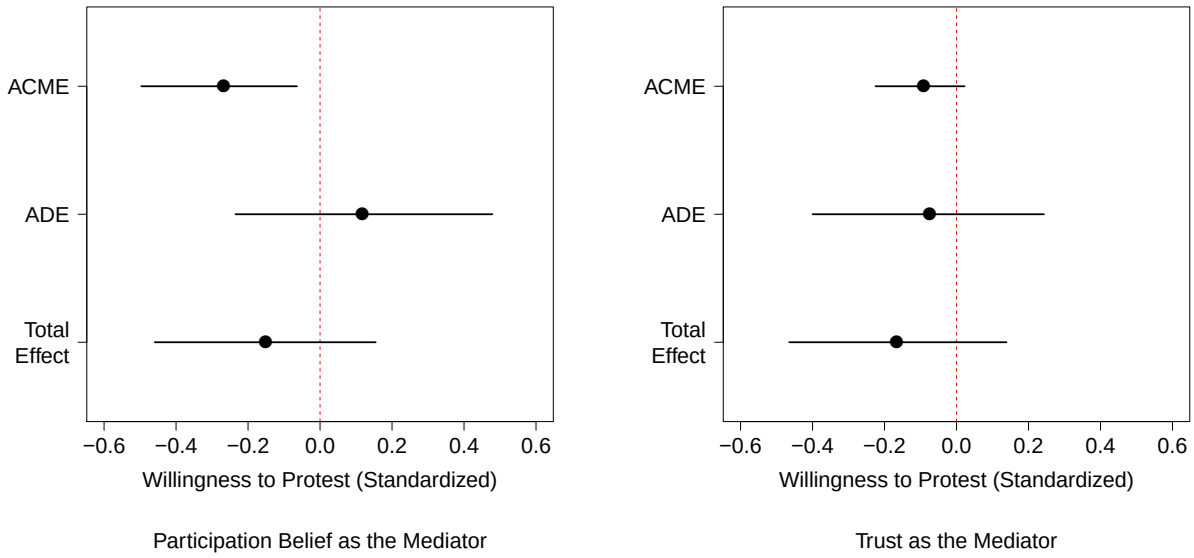


(Notes: 95% Confidence Intervals obtained via bootstrapping with 1000 resamples. See Online Appendix B.1 for the regression results underlying these figures.)

Figure 5: Mediators between Digital Surveillance and Protest Participation

Figure 6 presents the causal mediation analysis of in-person surveillance on willingness to protest. Although the total effect is not statistically significant, the average causal mediation effect of participation belief on protest participation is still negative and statistically

significant. The results from Figure 5 and Figure 6 suggest that surveillance discourages protest participation mainly through influencing individuals' strategic consideration about others' participation rather than their trust levels.[9]



(Notes: 95% Confidence Intervals obtained via bootstrapping with 1000 resamples. See Online Appendix B.1 for the regression results underlying these figures.)

Figure 6: Mediators between In-person Surveillance and Protest Participation

4.4 Discussion

Experimenter demand effects may bias estimates from a within-subject design because experimental subjects may tailor their responses to conform to their perceptions of the researcher's hypothesis due to answering the same question twice (before and after treatment). However, recent literature demonstrates that demand effects are typically modest and usually do not alter the treatment effects in survey and field experiments; further research participants often exhibit a limited ability to infer researchers' expectations (De Quidt, Haushofer

[9]Note that belief about others' turnout is different from generalized trust. An individual may trust others but still think the others would not participate under repression threat. The fact that participation belief serves as a mediator rather than trust suggests that the relationship between these two variables is weak.

and Roth 2018; Mummolo and Peterson 2019). In addition, experimenter demand effects would likely produce similar outcomes for both digital and in-person surveillance treatment groups since respondents in both groups are asked repeated questions in the same way. Thus, there would not have to be heterogeneous – across the digital and in-person treatments – demand biases to affect inference from a finding for differences between these two types of surveillance. In addition, because respondents complete the survey in less than 10 minutes, it is difficult for them to differentially infer the researcher's expected outcome from the two treatments.

I formally assess the potential demand effects by comparing post-treatment responses between the within-subjects design and the between-subjects design. Because demand effects are less likely to present in the between-subjects design that has no repeated questions, if there are any demand effects, we would expect systematic differences in post-treatment responses between these two designs. Online Appendix B.3 presents the details of the comparison, which shows that the differences in regime legitimacy, online expression, and protest participation between the two designs are statistically *insignificant*, and the differences in beliefs about others' participation are inconsistent between digital surveillance and in-person surveillance. Only trust exhibits systematic differences between the two designs. However, it should be noted that the differences may also be driven by respondents' heightened awareness of the treatments that we want to foster through the within-subjects design.

The spillover effect of the treatments is also unlikely to be a problem. First, the survey was not conducted in classrooms and dormitories where respondents would more often be classmates or roommates. The enumerators randomly approached individuals on the roads or in dinning halls and respondents were strangers to each other. Occasionally, respondents came in groups, but they did not communicate during the survey. This strategy reduces the likelihood of spillovers. Practically, if there were information spilling over from the treatment groups to the control group, such spillovers would downwardly bias estimates towards 0. Moreover, information spillovers between treatment groups would lead to similarities in the

effects of in-person and digital surveillance. For example, if the negative effect of digital surveillance on trust were due to information spilling over from the in-person surveillance treatment group, we would find a negative effect of digital surveillance on regime legitimacy similar to the in-person surveillance effect. Instead, we see significant difference between the two treatment groups. This further rules out information spillovers.

Finally, one might be concerned about the statistical inferences based on standard comparisons of means using OLS estimation. As an alternative, I use randomization inference methods to examine the statistical significance of the results (Athey and Imbens 2017). I randomly assign (fictional) treatment status and estimate treatment effects 1,000 times. I then compare the t-statistics from the estimated treatment effects from the fictional treatment assignments to the t-statistics from the actual treatment assignment. Two-sided tests find very similar p-values for the treatment effects of surveillance on trust, legitimacy, and protest participation. See Online Appendix B.5 for details.

5 Interrupted Time Series Design

To widen the temporal and spatial scopes of the experimental study, I provide additional evidence for the theory using a nationally representative sample of Chinese citizens. To be specific, I use the 2015 Chinese General Social Survey with a sample of 10,968 respondents and an interrupted time series design that exploits an exogenous shock to the Chinese government's digital surveillance practice caused by the 2015 Tianjin Explosions. This design focuses on the effects of digital surveillance on trust and civic participation. Due to the limitation of observational data, I am not able to compare digital surveillance with in-person surveillance under this setting. Nevertheless, recent empirical studies find that traditional, Stasi-style surveillance has long-lasting negative effects on interpersonal trust, institutional trust (i.e., regime legitimacy), and election participation, which provide external validity to my experimental findings on in-person surveillance (e.g., Lichter, Loeffler and Siegloch 2019).

5.1 2015 Tianjin Explosions and Government Surveillance

On August 12, 2015, a series of blasts in a Sinochem subsidiary's warehouse in the port of Tianjin killed 173 people and injured nearly a thousand (Merchant 2017). More than 17,000 households were damaged by the explosion, and 779 businesses suffered losses. The two major explosions were caused by combustible fertilizer ammonium nitrate, detonated by fire and small explosions due to the misuse of fire water sprinklers on some chemicals (Huang and Zhang 2015). According to the waveform record results, the first major explosion occurred at 11:34:06 pm, and the local earth-quake magnitude ML was about 2.3. The second major explosion occurred 30s later, and the local earthquake magnitude ML was about 2.9. The resulting fireballs reached hundreds of meters in height. The second explosion was estimated to be 336 tons TNT equivalent (Huang and Zhang 2015). Days later, local authorities ordered the evacuation of residents within a 3 km (1.9 mi) radius of the blast site, prompted by the threat of "toxic substances", including sodium cyanide (Ryan 2015).

Immediately after the blasts, information on the event, including blast videos, was released over social media sites and network platforms like Weibo and Wechat. This accident drew a great deal of attention among Chinese netizens, with the topic racking up more views on Weibo than the country's total population of nearly 1.4 billion (Dou 2015). Figure 7 shows the temporal distribution of Internet search activities in Mainland China measured by the Google search index using "Tianjin" as the keyword. We can see the search intensity peaked in the two weeks immediately following the accident.[10]

The devastating explosions raised serious questions about corruption, industrial safety, and emergency responses in China (Merchant 2017; Dou 2015). As soon as discussions and rumors went viral on the Internet and social media, the country's Internet surveillance and

[10]Although Google is blocked in Mainland China, people can use virtual private networks (VPNs) to circumvent the firewall. Thus, Google search index still reflects the temporal trend of public interests.

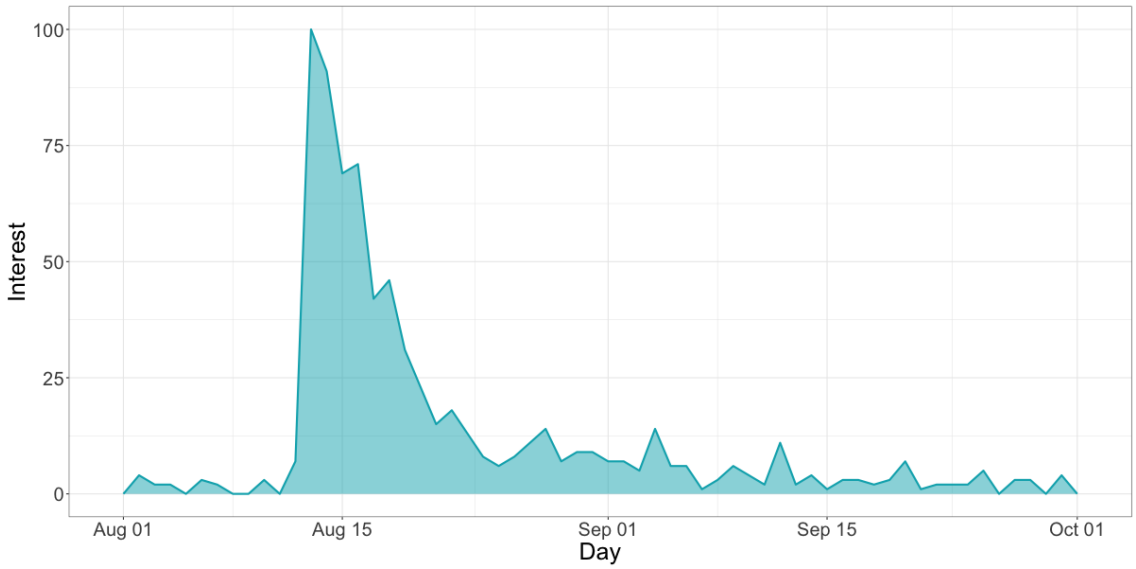


Figure 7: Google Trends on Tianjin Explosions

censorship machines operated at full capacity to control information and silence discussions. Data from censorship tracker Weiboscope, built by an associate professor at the University of Hong Kong's Department of Journalism, shows that surveillance and censorship rates on Weibo were up tenfold after the explosions compared to earlier in the month (Dou 2015). Large amounts of posts and discussions were deleted on the Internet and social media, including facts about the casualties, pictures and videos of the explosions and the site afterward, investigative reports, as well as comments that question the political ties of the warehouse owners, criticize the government's responses, and discuss the chemicals inside the warehouse (Dou 2015; Hanrahan 2015). Conceptually, censorship is the combination of online surveillance and repression – censors need to identify the objectionable posts first and then delete them. The Tianjin accident creates exogenous shock to censorship, which can be used to identify the effects of digital surveillance (and repression) on individuals' trust and civic participation.

5.2 Design and Data

My research design takes advantage of the coincidence that the 2015 Chinese General Social Survey (CGSS) was being conducted across China around the time of Tianjin blasts.

The CGSS is a nationwide survey ran every other year by the China Survey and Data Center at Rinming University – one of the top research universities in China. It participates in the International Social Survey Program and is the most reliable social survey in China. In Figure 8, I present the time, location, and number of individuals surveyed in the 2015 CGSS, with the dashed line indicating the time of Tianjin blasts.

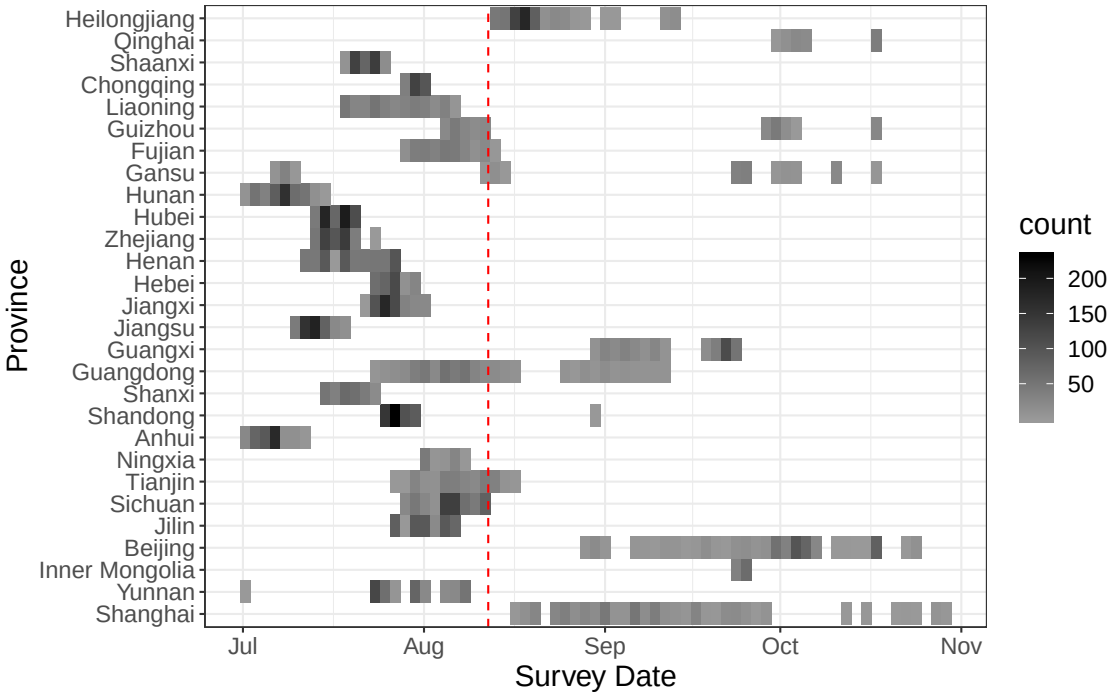


Figure 8: Distribution of Survey Respondents in the 2015 CGSS

Following Mummolo (2018), I use the interrupted time series (ITS) approach to compare individuals surveyed right before the Tianjin accident with individuals surveyed right after it.[11] As shown in Figure 8, some provinces do not have observations either before or after the accident. Because China is a large country with 31 provinces where people may hold different beliefs, the ITS estimate could be biased by regional differences. Thus, I include province fixed effects into the model. Specifically, I estimate the following equation:[12]

[11]Scholars also refer this approach to a regression discontinuity in time design (e.g., Hausman and Rapson 2018).

[12]The nonparametric method with robust confidence interval developed by Calonico, Cat-

$$Y_{ip} = \alpha + \delta\, cutoff_{ip} + \pi\, min_{ip} + \lambda\, cutoff_{ip} \cdot min_{ip} + X'_{ip}\,\Psi + province_p + \epsilon_{ip} \qquad (1)$$

where i indexes the respondent and p the province; $cutoff_{ip}$ is a binary variable that takes the value of 1 if the respondent was interviewed after 11:34:06 pm on August 12, 2015 and 0 otherwise; min_{ip} is the running variable – the minute when the respondent was interviewed; X'_{ip} is a set of individual controls; and $province_p$ is the province fixed effects. In this specification, I use a liner function to model time trend on either side of the cutoff because peoples' beliefs often do not change drastically. Nevertheless, online Appendix C.2 show that the results are largely robust to models with quadratic and cubic function forms.

For the dependent variables, Y_{ip}, I use four measures: interpersonal trust, regime legitimacy, views about expression, and views on petitioning. Table 5 shows the questions for these measures from the CGSS.

Table 5: Measures of Trust, Legitimacy, and Participation in the CGSS

View on Free Speech: To what extent does the following statement reflect the reality in China? The right to criticize the government publicly is protected by law.
View on Petition: Do you agree or disagree with the following statement concerning petitions in China? Petitions are not obstructed.
Interpersonal Trust: Generally speaking, do you agree that most people can be trusted in the society?
Regime Legitimacy: Do you agree with the following statement? Some policy reforms of the government bodies violate current laws but those policies have good intentions and work well. Such government reforms deserve recognition and appraisal.

This empirical strategy identifies the effect of digital surveillance that may not apply to individuals who were not exposed to the intensified online surveillance after the Tianjin accident. I then use individuals' Internet usage (including WAP phone services) to identify and exclude those who do not use the Internet (approximately half of the sample).[13] In

taneo and Titiunik (2014) is not suitable for fixed-effects models. Thus, I use a parametric method for estimation.

[13]After sample restriction, it is still much more representative than the student sample in

addition, the identification assumption of the ITS approach requires the groups of individuals surveyed before and after the time cutoff to be identical. As other concurrent events might influence respondents surveyed much earlier or much later, I further restrict the samples to one-week, two-week, and three-week windows (i.e., one week before and one week after, so on and so forth). I use the two-week window sample for the main analysis because people's interests peaked within two weeks after the Tianjin accident (Figure 7). The different windows also serve as robustness checks of the ITS estimates. Appendix C.3 shows the summary statistics of the samples.

The ITS design could be weakened if there was "precise" sorting of the survey respondents around the cutoff time to imbalance the treated and control groups. This is unlikely the case since, first, respondents could not decide when they were interviewed, and, second, survey organizers could not change the predetermined sampling scheme to select a biased sample after the accident. In addition, survey interviewers all over the country had neither incentives nor capabilities to systematically select a different aftermath sample specifically biased in terms of trust and participation. I further examine the conditions of local randomization by looking at whether baseline covariates are balanced. Online Appendix C.2 shows that a number of covariates have no statistically significant changes around the cutoff time.[14]

5.3 Design for Mechanism Testing

One concern is whether the effects identified by the ITS approach are due to surveillance caused by the Tianjin explosions or just due to the accident itself. Other unknown concurrent events could also bias the results. To examine the mechanism of digital surveillance, I

the survey experiment.

[14]I check for discontinuities in other covariates at the threshold instead of using the McCrary density test to examine the sorting problem because the density of the running variable (time) is uniform, which renders the test for discontinuities in its conditional density irrelevant (Hausman and Rapson 2018).

construct a measure of provincial-level surveillance intensity using the number of pilot counties for China's Golden Shield Project (GSP) – a domestic digital surveillance and content filtering system that integrates on online government databases with an all-encompassing surveillance network developed by the Ministry of Public Security (Walton 2001).

The phase-in GSP was implemented in small scale in some prefectures in 2000s but carried over in large scale in early 2010s, especially the "3111" Initiative that built local networks of digital surveillance tools with integrated street surveillance cameras. Under the "3111" Initiative, a total of 660 pilot counties/districts were selected by provincial governments between 2008 and 2012 to build the surveillance camera systems. The ratio of pilot counties to the total counties in a province reflects the strength of digital surveillance in that province because the operation of surveillance camera systems requires well-developed surveillance infrastructures, integrated surveillance platforms, and sufficient security personnel. In addition, although digital surveillance and censorship were mainly operated online, local security force is required to enforce punishment such as intimidation, harassment, and detention (Mozur 2019). Thus, this ratio reflects the intensity of surveillance and repression in a province.

Following Fitzpatrick (2010), I include an interaction term into Equation (1) to examine how the effects of surveillance and censorship caused by the Tianjin accident vary at different level of surveillance intensity. This is an extension of the above hypotheses since it assumes the effects of digital surveillance on civil participation are conditional on the intensity of surveillance and repression. In particular, I estimate the following specification.

$$Y_{ip} = \alpha + \delta\, cutoff_{ip} + \pi\, min_{ip} + \beta\, intensity_p + \gamma\, cutoff_{ip} \cdot intensity_p + \lambda\, cutoff_{ip} \cdot min_{ip} + X'_{ip}\,\Psi + province_p + \epsilon_{ip} \quad (2)$$

where $intensity_p$ is the ratio of "3111" pilot counties to the total counties in a province and $cutoff_{ip} \cdot intensity_p$ the interaction term. I expect γ to be positive and statistically significant.

6 Findings from the 2015 CGSS Sample

This section presents the results of the Interrupted Time Series design using the 2015 CGSS data. I first present the main effects of the Tianjin explosions on trust and civic participation. Then, I show how the effects change with the level of surveillance intensity.

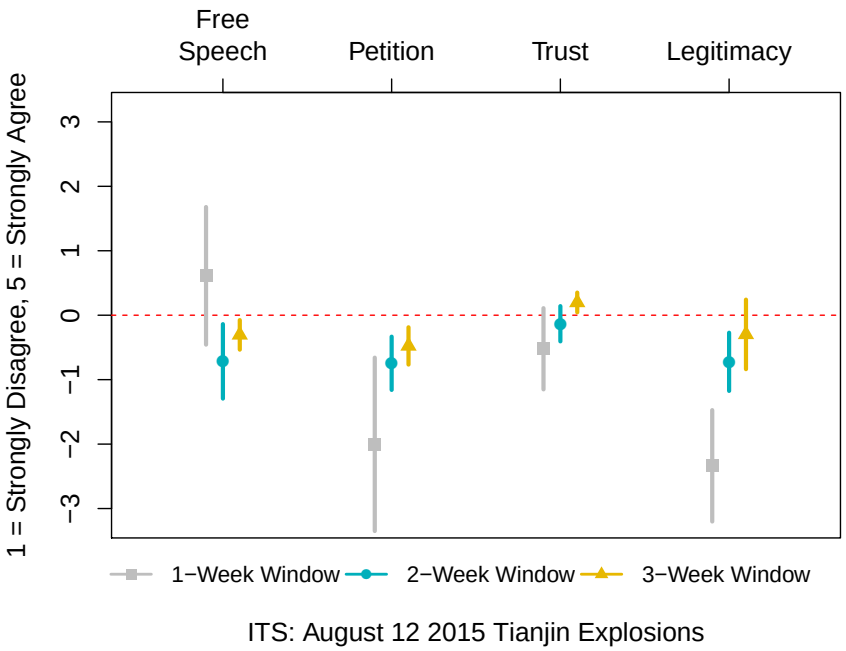
6.1 Main Effects

Figure 9 plots the effects of digital surveillance (caused by the Tianjin explosions) on individuals' view of free speech, perceived risk of petition, interpersonal trust, and regime support. Under intensified digital surveillance, individuals are less likely to think that the right to criticize the government publicly is protected by law. They are more likely to think that petitions are obstructed by the government. The effect on view of petition holds for all the samples within one-week, two-week, and three-week windows. The effect on view of free speech hold for the two-week and three-week windows. On the other hand, the effect of digital surveillance on trust is negative but statistically insignificant. These findings are consistent with the experimental results.

The theory and the experimental findings suggest that digital surveillance should not reduce regime legitimacy. However, we find a negative effect of the Tianjin accident on regime legitimacy in the one-week and two-week windows. This inconsistency is likely due to the negative effect of the accident itself: after the horrifying explosions, people blamed the government for its failures to contain corruption, ensure industrial safety, and respond to emergencies, which, in turn, lower regime support.

6.2 Mechanism Testing

I further examine whether the negative effects of Tianjin explosions on individuals' beliefs are conditional on the intensity of digital surveillance. In Figure 10, I plot the results of the ITS model with interaction between the time cutoff and surveillance intensity. The left panel shows that the marginal effect of Tianjin explosions on individuals' view of free speech

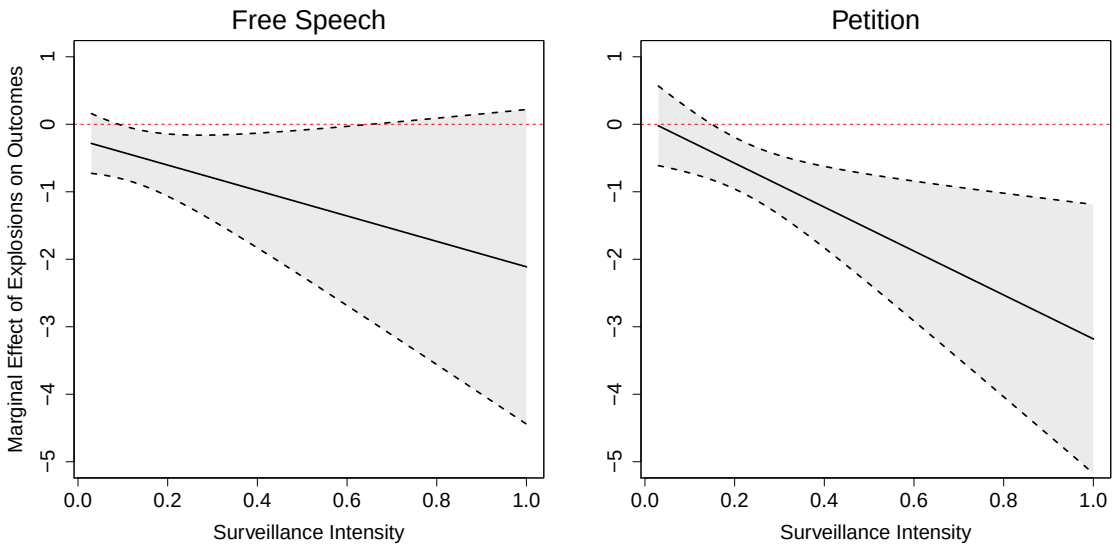


(Notes: OLS estimates with 95% Confidence Intervals. Standard errors clustered at the prefecture level. See Online Appendix C.3 for the regression results underlying these figures.)

Figure 9: Tianjin Explosions on Civic Participation, Trust, and Legitimacy

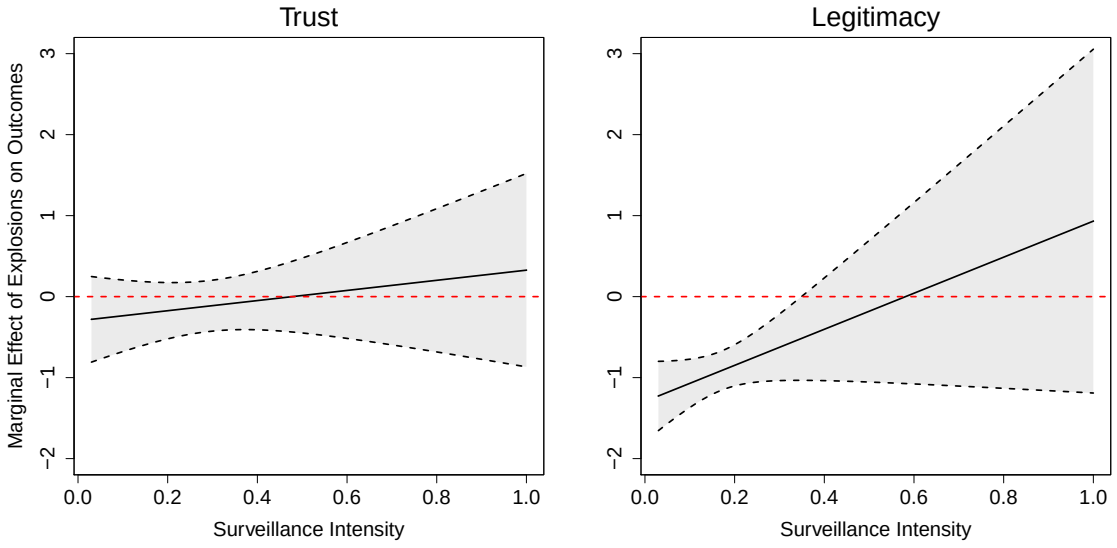
decreases with higher surveillance intensity and the effect is statistically significant when the intensity is within the 0.1 – 0.7 range. The right panel shows that the marginal effect of explosions on view of petition also decreases with higher surveillance intensity and is statistically significant at most of surveillance intensity levels. These findings provide strong evidence that digital surveillance is the driving force behind the decreased civic engagement after the Tianjin explosions.

Figure 11 plots the marginal effects of Tianjin explosions on interpersonal trust and regime legitimacy. Consistent with our expectation, the marginal effect on trust is statistically insignificant. Interestingly, the marginal effect of explosions on regime legitimacy increases with higher surveillance intensity and is statistically significant when the intensity is within the 0 – 0.4 range. This provides further evidence that the reduced legitimacy is due to the Tianjin accident itself because stronger surveillance and censorship prevent the information about the accident from spreading to lower citizens' regime support.



(Notes: OLS estimates with 95% Confidence Intervals. Standard errors clustered at the prefecture level. See Online Appendix C.3 for the regression results underlying these figures.)

Figure 10: Marginal Effects of Explosions on Civic Participation



(Notes: OLS estimates with 95% Confidence Intervals. Standard errors clustered at prefecture level. See Online Appendix C.3 for the regression results underlying these figures.)

Figure 11: Marginal Effects of Explosions on Trust and Legitimacy

The above findings from the 2015 CGSS sample are largely consistent with the experimental findings, which lend further weight to the main theoretical claims.

6.3 Discussion

Some concurrent events could potentially bias the estimates of the ITS approach. Fortunately, during the four-week window there were no big events or incidents that could invalid the surveillance treatment caused by the Tianjin Explosions. On July 26, an escalator death of a 31 year-old mother set off furor online, but this incident did not incur surveillance and censorship. Even if there were censorship, it would lead to underestimation because censorship on this incident occurred before the Tianjin explosions. There were a few small incidents such as a chemical explosion in Shandong (August 31, 5 death) and landslide in Shaanxi (August 12, 26 missing), but none of them caught nation-wide attention in China. Two relatively large events or incidents are stock market rollercoaster (June 12 – July 10) and China's Victory over Japan Day 70th Anniversary Parade (September 3), but both events were not in the four-week treatment window. Even if they had influence on citizens during the four-week window, the stock market clash would have reduced trust and regime legitimacy before the treatment time, and the anniversary would have increased regime legitimacy after the treatment time. Thus, both events would lead to underestimation of the treatment effects.

Another concern is that the results could be driven by the differences between individuals surveyed before and after the Tianjin incident. Individuals from different regions may hold different opinions to bias the ITS estimates if the pre- and post-treatment samples were drawn from difference regions. I control for province fixed effects to reduce this potential bias. Further, I examine the effects of the Tianjin incident on *non-netizens'* opinions. The logic is that, if regional differences drove the treatment effects, we would see similar patterns on non-netizens who did not received the surveillance treatment. However, I find no difference between the pre- and post-treatment groups in the non-netizen sample (Appendix B.3). In addition, the Tianjin incident could cause some new people to become Internet users afterwards, changing the post-treatment sample. Although I cannot rule out this possibility, the fact that the covariates are balanced between the pre- and post-treatment groups suggests

self-selection is not a concern.

Moreover, the mechanism test that the Tianjin Explosions event has stronger effects in areas with more intensive surveillance further suggests that concurrent events and sample selection are unlikely to invalid the findings.

7 Conclusion

Dictatorships have a long history of using surveillance for information collection and social control. Traditional, in-person surveillance was the main form until the onset of the information age. While the free flow of digital information has had transformative effects on economic well-being for citizens throughout the world, it is increasingly important to understand how dictatorial governments exploit this information for surveillance, and ultimately political control. This paper is the first to compare the social costs of digital surveillance with those of in-person surveillance. Using an in-the-field experiment in China, I show that both types of surveillance deter expression and protest participation, but digital surveillance is less likely to lower interpersonal trust and regime legitimacy. Evidence from a natural experiment using the 2015 Chinese General Social Survey is consistent with the empirical findings. The theory and results imply that digital surveillance is a very efficient tool of authoritarian control: it deters expression and protest participation and, meanwhile, does not yield the same potential costs for dictators that human surveillance entails.

The empirical findings in this paper are based on individuals' short-term attitudes toward surveillance operations in China, but the theory and implications travel to long-run civic culture in other authoritarian countries. For example, Lichter, Loeffler and Siegloch (2019) combine a border discontinuity design with an instrumental variables approach to study the long-run effect of the Stasi's in-person surveillance on civic capital and economic performance in Germany. They find that a larger share of county-level informants in the population in 1980s led to persistently lower levels of civic participation, interpersonal trust, and regime support in post-reunification Germany. They also find substantial, long-lasting, negative

effects of in-person surveillance on economic outcomes. In Iraq, low levels of social trust and political expression appear in places under heavier state surveillance in Iraq (Blaydes 2018). My findings on in-person surveillance are largely consistent with previous work. Although we have limited evidence on the long-term effects of digital surveillance on trust and regime legitimacy, the findings that digital surveillance deters expression and protest coordination should hold in the long run given its strong deterrence effects even in the survey experiment.

This paper compares the social costs of digital surveillance with in-person surveillance. Yet, this is not to say dictators prefer one tool to the other. Evidently, authoritarian governments use both tools together to control society. Many websites and social media platforms in China, for example, provide shortcut links and encourage users to report suspicious posts to the state authority. Recently, there is a growing trend of reporting political comments on China's social media (Yan 2019). Moreover, these two types of surveillance can serve different purposes. Digital surveillance is quicker and more comprehensive than human-agent based surveillance. It is useful to screen out suspicious opponents from a very large population. Once the targets are identified, dictators may rely on human agents and informants for more detailed information collection. As dictators expand their information-collection toolkit with advanced digital surveillance tools, civil virtue is likely to decline further in authoritarian countries in the digital age.

References

Algan, Yann and Pierre Cahuc. 2010. "Inherited Trust and Growth." *American Economic Review* 100(5):2060–92.

Almond, Gabriel Abraham and Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton university press.

Ash, Timothy Garton. 1998. *The File: a Personal History*. Vintage.

Athey, Susan and Guido W Imbens. 2017. The Econometrics of Randomized Experiments. In *Handbook of Economic Field Experiments*. Vol. 1 Elsevier pp. 73–140.

Benson, Michelle and Thomas R Rochon. 2004. "Interpersonal Trust and the Magnitude of Protest: a Micro and Macro Level Approach." *Comparative Political Studies* 37(4):435–457.

Blaydes, Lisa. 2018. *State of Repression: Iraq Under Saddam Hussein*. Princeton University Press.

Bruce, Gary. 2010. *The Firm: the Inside Story of the Stasi*. Oxford University Press.

Calonico, Sebastian, Matias D Cattaneo and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-discontinuity Designs." *Econometrica* 82(6):2295–2326.

Cantoni, Davide, David Y Yang, Noam Yuchtman and Y Jane Zhang. 2019. "Protests as Strategic Games: Experimental Evidence From Hong Kong's Antiauthoritarian Movement." *The Quarterly Journal of Economics* 134(2):1021–1077.

Chwe, Michael Suk-Young. 2013. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton University Press.

De Quidt, Jonathan, Johannes Haushofer and Christopher Roth. 2018. "Measuring and Bounding Experimenter Demand." *American Economic Review* 108(11):3266–3302.

Deibert, Ron. 2017. "Evidence That Ethiopia is Spying on Journalists Shows Commercial Spyware is Out of Control." *Wired* . December 06.

Dimitrov, Martin K. and Joseph Sassoon. 2014. "State Security, Information, and Repression:

A Comparison of Communist Bulgaria and Ba'thist Iraq." *Journal of Cold War Studies* 16(2):3–31.

Dou, Eva. 2015. "China's Censors Scramble to Contain Online Fallout After Tianjin Blast." *The Wall Street Journal* . August 16.

Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *Review of Economic Studies* 80(4):1422–1458.

Faust, Aaron M. 2015. *The Ba'thification of Iraq: Saddam Hussein's Totalitarianism*. University of Texas Press.

Fitzpatrick, Maria Donovan. 2010. "Preschoolers Enrolled and Mothers at Work? The Effects of Universal Prekindergarten." *Journal of Labor Economics* 28(1):51–85.

Fukuyama, Francis. 1995. "Social Capital and the Global Economy." *Foreign Aff.* 74:89.

Gieseke, Jens. 2014. *The History of the Stasi: East Germany's Secret Police, 1945-1990*. Berghahn Books.

Greif, Avner. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." *The journal of economic history* 49(4):857–882.

Greitens, Sheena Chestnut. 2016. *Dictators and Their Secret Police: Coercive Institutions and State Violence*. Cambridge University Press.

Gunitsky, Seva. 2015. "Corrupting the Cyber-commons: Social Media as a Tool of Autocratic Stability." *Perspectives on Politics* 13(1):42–54.

Hanrahan, Mark. 2015. "Tianjin Explosions: Disasters In China Prompt Wave Of Media Censorship." *International Business Times* . August 13.

Hausman, Catherine and David S Rapson. 2018. "Regression Discontinuity in Time: Considerations for Empirical Applications." *Annual Review of Resource Economics* 10:533–552.

Huang, Ping and Jingyuan Zhang. 2015. "Facts Related to August 12, 2015 Explosion Accident in Tianjin, China." *Process Safety Progress* 34(4):313–314.

Imai, Kosuke, Luke Keele and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." *Psychological methods* 15(4):309.

Intel, Public Relations. 2017. "Intel Movidius Helps Bring Artificial Intelligence to Video Surveillance Cameras." *Intel Newsroom* . April 5. Available at: https://newsroom.intel.com/news/intel-movidius-helps-bring-artificial-intelligence-video-surveillance-cameras/#gs.unl5w3. Accessed August 9, 2019.

Kalyvas, Stathis N. 2006. *The Logic of Violence in Civil War*. Cambridge University Press.

Kendall-Taylor, Andrea and Erica Frantz. 2014. "How Autocracies Fall." *The Washington Quarterly* 37(1):35–47.

King, Gary, Jennifer Pan and Margaret E. Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107(2):326–343.

King, Gary, Jennifer Pan and Margaret E. Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." *American Political Science Review* 111(3):484–501.

Knabe, Hubertus. 1999. "Die feinen Waffen der Diktatur. Nicht-strafrechtliche Formen politischer Verfolgung in der DDR." *Die DDR–Erinnerungen an einen untergegangenen Staat* 88:191–220.

Krosnick, Jon A and Leandre R Fabrigar. 1997. "Designing Rating Scales for Effective Measurement in Surveys." *Survey measurement and process quality* pp. 141–164.

Kuran, Timur. 1991. "Now Out of Never: The Element of Surprise in the East European Revolution of 1989." *World Politics* 44(1):7–48.

Lichter, Andreas, Max Loeffler and Sebastian Siegloch. 2019. "The Economic Costs of Mass Surveillance: Insights From Stasi Spying in East Germany." Working Paper. Available at: https://www.dropbox.com/s/dm253x8bh8nspgo/Stasi_onlineversion_aug2019.pdf?dl=0. Accessed August 3, 2019.

Liu, Joyce and Xiqing Wang. 2017. "In Your Face: China's all-seeing state." *BBC News* . December 10.

Lorentzen, Peter L. 2013. "Regularizing Rioting: Permitting Public Protest in an Authori-

tarian Regime." *Quarterly Journal of Political Science* 8(2):127–158.

Lundmark, Sebastian, Mikael Gilljam and Stefan Dahlberg. 2015. "Measuring Generalized Trust: An Examination of Question Wording and the Number of Scale Points." *Public Opinion Quarterly* 80(1):26–43.

McMillan, John and Pablo Zoido. 2004. "How to Subvert Democracy: Montesinos in Peru." *Journal of Economic perspectives* 18(4):69–92.

Merchant, Nomann. 2017. "China Investigates Former Local Party Boss for Bribery." *The Associated Press News* . January 22.

Moore, Kelly. 1999. "Political Protest and Institutional Change: the Anti-vietnam War Movement and American Science." *How social movements matter* 10:97–118.

Mozur, Paul. 2019. "Twitter Users in China Face Detention and Threats in New Beijing Crackdown." *New York Times* . January 10.

Mummolo, Jonathan. 2018. "Modern Police Tactics, Police-citizen Interactions, and the Prospects for Reform." *The Journal of Politics* 80(1):1–15.

Mummolo, Jonathan and Erik Peterson. 2019. "Demand Effects in Survey Experiments: an Empirical Assessment." *American Political Science Review* 113(2):517–529.

Putnam, Robert D, Robert Leonardi and Raffaella Y Nanetti. 1994. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton university press.

Qin, Bei, David Strömberg and Yanhui Wu. 2017. "Why Does China Allow Freer Social Media? Protests Versus Surveillance and Propaganda." *Journal of Economic Perspectives* 31(1):117–40.

Ritter, Emily Hencken and Courtenay R. Conrad. 2016. "Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression." *American Political Science Review* 110(1):85–99.

Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall*. Princeton University Press.

Ryan, Fergus. 2015. "China Explosions: Police Order Mass Evacuations Amid Further

Blasts." *The Guardian* . August 15.

Singh, Amarjot, Devendra Patil, Meghana Reddy and SN Omkar. 2017. Disguised Face Identification (DFI) With Facial Keypoints Using Spatial Fusion Convolutional Network. In *Proceedings of the IEEE International Conference on Computer Vision*. pp. 1648–1655.

Sullivan, Christopher M. 2016. "Political Repression and the Destruction of Dissident Organizations: Evidence from the Archives of the Guatemalan National Police." *World Politics* 68(4):645–676.

Valentino-DeVries, Jennifer, Lam Thuy Vo and Danny Yadron. 2015. "Cataloging the World's Cyberforces." *The Wall Street Journal* . December 28.

Walton, Greg. 2001. *China's Golden Shield: Corporations and the Development of Surveillance Technology in the People's Republic of China*. Rights & Democracy.

Wintrobe, Ronald. 2000. *The Political Economy of Dictatorship*. Cambridge University Press.

Wiswall, Matthew and Basit Zafar. 2014. "Determinants of College Major Choice: Identification Using an Information Experiment." *The Review of Economic Studies* 82(2):791–824.

Xu, Xu and Xin Jin. 2018. "The Autocratic Roots of Social Distrust." *Journal of Comparative Economics* 46(1):362–380.

Yan, Alice. 2019. "Chinese Teacher Suspended for 'Belittling' Great Inventions in Online Chat." *South China Morning Post* . August 26. Available at: https://www.scmp.com/news/china/society/article/3024372/chinese-teacher-suspended-belittling-great-inventions-online. Accessed September 3, 2019.

Zhao, Dingxin. 2004. *The Power of Tiananmen: State-society Relations and the 1989 Beijing Student Movement*. University of Chicago Press.

Zuo, Qiuming. 1998. *Discourses of the State (BC 500)*. Shanghai Ancient Books Publishing House.