GÖTEBORGS UNIVERSITET

DEPARTMENT OF POLITICAL SCIENCE

Trust is good; control is better

Exploring repression in the relation between Collective Actions and Blacklists within the Chinese Social Credit System

Alexandre Souza Gomes

Master's Thesis:	30 higher education credits
Programme:	Master's Programme in Political Science
Date:	08/2020
Supervisor:	Seraphine Maerz
Words:	15901

Table of contents

List of figures and tables

Figures

Tables

Acronyms and abbreviations

AI	Artificial Intelligence
CA	Collective Action
CASM	Collective Action from Social Media
CCP	Chinese Communist Party
CCP	Chinese Communist Party
COV	Co-variational analysis
CDB	Court Defaulter Blacklist
CV	Control Variable
DV	Dependent Variable
ICT	Information and Communication Technologies
IV	Independent variable
GDP	Gross Domestic Product
GDPpc	Gross Domestic Product per capita
HCA	High Capacity Autocracies
ICT	Information and communication Technologies
OECD	Organisation for Economic Co-operation and Development
PBOC	People's Bank of China
RMB	Renminbi (yuan) - China's currency
SCS	Social Credit System
SPC	Supreme People's Courts
VCA	Violent Collective Action
VCApc	Violent Collective Action per capita

Abstract:

The Social Credit System (SCS) is a trial regulatory and reputational system set to score, reward, and punish Chinese residents for desirable and undesirable behavior. Officially, the SCS aims at enhancing overall societal trust, and integrity. The autocracy literature takes issue with its repressive potential to surveil and control society but lacks both cohesive theoretical frameworks and empirical evidence to explore it. To fill this gap, I addressed how the SCS could strengthen the stability of the Chinese regime by enhancing repression, legitimation, and co-optation. Focusing on repression, I examined how the most advanced part of the SCS, the Court Defaulter Blacklist (CDB), can be considered a new form of non-violent repression in response to Collective Actions, which results to the following research question: *do more Violent Collective Actions (X) lead to more Court Defaulter Blacklist (Y)?* To address this, I used a cross-sectional co-variational analysis case study on the county level. The case selection was based on Zhang and Pan's (2019) Collective Action from Social Media dataset throughout 1162 counties from 2010 to 2017 (X) followed by other relevant control variables. The CDB data was independently collected from county court performance reports (2015-2017). The final selection had 3 counties with high and 3 with low incidence of Violent Collective Actions (VCAs), with the former having 5 times more CDBs than the latter. This evidence confirms the effects between VCA and the CDB, further backing claims that the SCS is used to repress, casting doubt on the SCS's official rhetoric, and serving as a plausibility probe for potential large-N analysis.

Keywords: Chinese social credit system, autocracy, repression, Collective Action, blacklist

Chapter 1: Introduction

"An Orwellian system premised on controlling virtually every facet of human life" (Pence 2018), and "frightening and abhorrent structure" (Soros 2019). Those are only two of the many headlines about the Chinese Social Credit System (SCS) launched in 2015. Technically, this is a big data pilot system built to score, reward, and punish all adult Chinese residents' behavior based on a massive pool of financial and non-financial behavior that, at least according to the Chinese Communist Party (CCP), aims at building a more trustworthy society (Ohlberg, Ahmed, and Lang 2017, 6). Regardless of this contradiction, when fully implemented, this system could affect up to ⅙ of the world's population, and inspire others to follow, hence its importance.

Researchers in different fields have done a good job describing the system, drawing comparisons between different social credit and rating systems worldwide (Mac Síthigh and Siems 2019). Others theorized how it could be a tool for social control (Botsman 2017; Falkvinge 2015). There is also work pointing to the exaggerated surveillance dangers of building such a massive big data structure (Liang et al. 2018; Mosher 2019; Qiang 2019). However, only a handful have investigated the SCS quantitatively, namely its positive approval ratings (Kostka and Antoine 2018), how it is enthusiastically communicated to the public (Ohlberg, Ahmed, and Lang 2017), and how specific SCS' components successfully frame bad behavior and its punishments but fail in individualizing what good behavior means, and how it should be rewarded (Engelmann et al. 2019).

In a broader context, autocracy researchers have explored the incentives and the forms to which autocrats seek long-term stability, particularly regarding different uses of repressive means (Gerschewski 2013; Levitsky and Way 2002; Dukalskis and Gerschewski 2018). As a response to protests and dissidence, the contentious politics literature also points to an autocratic trend to look for new, and non-violent forms of repression (Goldstone and Tilly 2001; Brumberg 2002; Gurr 1986). To illustrate this, recent literature reviewed Information and Communication Technologies (ICTs) measures to coerce the population, like targeted internet access shutdowns, censoring undesirable online content, as well as the systematic deletion of critical social media posts (Hassanpour 2014; Gohdes 2015; King, Pan, and Roberts 2013; Qin 2017).

Nonetheless, given the novelty of the SCS, autocracy literature about this specific system is virtually nonexistent. Hence, researchers have not yet started debating how the SCS may be conceptualized within an autocratic logic. Particularly, existing articles have not examined nor tested how the SCS might influence autocratic stability, or if the claims about the SCS' repressive nature are true. To fill this gap, this study builds on previous literature to theorize how the SCS could enhance autocratic stability by strengthening the CCP's repression, legitimation, and co-optation mechanisms. It centers on repression to examine how the SCS' Court Defaulter Blacklist could be considered a new form of non-violent coercion in response to Collective Actions[1]. Altogether, it aims at addressing the following research question: do more Violent Collective Actions lead to more Court Defaulter Blacklists (CDB)?

On the one hand, the data for the independent variable (IV) and for the control variables (CV) is relatively accessible. For IV Collective Actions (X) I will use Zhang and Pan (2019)'s Collective Action from Social Media (CASM) dataset with 14524 violent events throughout 1162 counties from 2010 to 2017. Whereas GDP, population, urbanization, and internet penetration will be used as control variables (CVs). On the other hand, there is no database available for the CDB data (Dependent Variable - DV), and building my own large-N database would require too many resources for data collection, and coding at the county-level. To address this, I will use a small-N cross-sectional co-variational analysis (COV) with a rigorous case selection harnessing the IV and CV datasets to allocate time and resources comparing only the cases that would broaden the validity of the results the most. The CDB data will be independently, and manually mined from performance reports from county courts between 2015-2017 for the 6 counties selected.

This thesis is divided into four chapters, the first starts with a review of the existent literature, followed by some basic facts about the SCS and the Blacklists. Next, it articulates the relation between the SCS and regime stability, and why CDB can be considered as a form of repression under the contentious politics lens. The second chapter presents a methodological framework, starting by a summary of the theoretical argument, the research question, and the hypothesis, followed by a general explanation about the research design chosen, then it elaborated on the details regarding the chosen data, and it ends with the operationalization of the case selection. The fourth chapter analyzes the final case selection, followed by both an internal and an external validity discussion about the result. Subsequently, the most important findings and limitations of the work are presented, Lastly, the concluding remarks will briefly revisit all the aspects of the thesis, and finalize by framing its broader implications and importance.

All in all, this research aims to offer empirical evidence to verify claims that the SCS' coercive and controlling nature. It intends to fit the SCS within an autocracy theoretical framework for analyzing non-democratic stabilization. Empirically, it intends to serve as a plausibility probe for potential large-N analysis investigating the SCS' repressive aspects, as well as to examine if the official CCP's rhetoric about an SCS' aiming for trust and financial compliance applies.

[1]Collective Action and protests are similar terms that will be used exchangeably until the end of this chapter.

Chapter 2: Theoretical framework

2.1. Literature review

Because the system is fairly new, specialized literature is not abundant, particularly within the autocracy literature. However, despite the inaccuracies and controversies in the media regarding the capabilities of the Chinese SCS, there have been some studies (Botsman 2017; Falkvinge 2015; Liang et al. 2018; Mosher 2019; Qiang 2019) addressing the system since its announcement in 2014. Generally, scholars caution about the potential implications of such large big data structures in a notorious autocratic country like China.

On a neutral note, (Mac Síthigh and Siems 2019) put the Chinese SCS into perspective by comparing different social credit and rating systems operating across the world, including those in the West. They analyze the SCS' level of intervention and its effect on individuals and conclude that, from a legal point of view, the SCS goes much further in terms of general scope and enforcement capabilities. Other authors believe that the Chinese SCS is a tool for social and/or political control that operates within a "state surveillance infrastructure", as defined by Liang (2018, 12) or an "evolving practice of control" (Hongri Zhang 2017). Mckenzie and Meissner (2016, 52) defined it as "an all-encompassing system penetrating, controlling and shaping society".

While debating China's automated social management development, Samantha Hoffman[2] asserts that the nature of the SCS's functioning could be made political so that the country's social and economic development will be inseparable from the Communist Party's control (Gan 2019, 10). This would mean "the technological marriage of individual "responsibility" mechanisms and social control methodologies" (Hoffman 2017, 24). There is also debate on the use of big data-based surveillance, which allows states to track "everything about everyone at all times" (Andrejevic and Gates 2014: 190) to predict undesirable actions and behaviors and control them while increasing social and political activities as well (Shorey and Howard 2016).

Most of those claims highlight the system's damaging potential. However, they have not gone empirically far enough to test, if the SCS is indeed guilty of all those charges. There are only a very limited number of studies that take a quantitative approach to investigate the SCS at all. The first is Ohlberg, Ahmed, and Lang (2017)'s collection of over 60.000 articles from the news, official sources, social media, and blog forums, and bulletins about the SCS to capture how the system is being communicated to the Chinese population. Their results indicate the official intent to create a "cure-all solution" for a multitude of societal problems, they also identified many

[2] Visiting fellow at the Australian Strategic Policy Institute Cyber Center.

described objectives and also highlighted the fact that most of the population still does not know much about it by then (Ohlberg, Ahmed, and Lang 2017, 1).

Second, Kostka and Antoine (2018) conducted a cross-regional survey and interviews with SCS participants to explore how the SCS is being perceived by the population. The researchers found that the system enjoys very high approval ratings among the population, and that many citizens have changed their behavior because of the SCS. Last, Engelmann et al. (2019) gathered close to 200.000 entries of different blacklists systems to explore how the system perceives good and bad behavior. Their design focused on Beijing and found that the system is fairly clear on what presents clear Sanctions but is very vague when it comes to rewards.

All in all, particularly regarding the autocracy literature, it lacks a more comprehensive theoretical approach to understand how the SCS might influence a non-democratic context like China. Additionally, it also lacks empirical evidence to support the claims about the system being used to repress. To address this, I tackle how the SCS could strengthen repression, legitimation, co-optation, and hence overall regime stability. Focusing on the former, I examine how its most advanced component, the Court Defaulter Blacklist (CDB), can be considered a new form of non-violent repression, particularly to Collective Action, and asking the following research question: Do more Violent Collective Actions (X) lead to more Court Defaulter Blacklist (Y)?

2.2. China's Social Credit System, and the Court Defaulter Blacklist

This part outlines the SCS' assumed objectives, the big data structure behind it, and the reward, and punishment systems in place, particularly the CDB.

The Social Credit System (SCS) in China is a pilot big data regulatory and reputational system set to score, reward, and punish Chinese citizens for not behaving with integrity as determined by the CCP'. It started trials in 2009 before the onset of its 6-year pilot phase in 2014, following the State Council Notice regarding the launch of the "Planning Outline for the Construction of a Social Credit System (2014-2020)"(Schaefer and Yin 2019, 22–23).

The Outline shows that most of China's social issues derive from the country's overall lack of trust and punishment to eventual rule breakers (Liu 2019, 22). Hence, to address this problem, the CCP claims to have created the SCS aims to enhance overall societal trust, and social integrity. As per the official State Council's document wording, the SCS was created to *"strengthen sincerity in government affairs, commercial sincerity, social sincerity, and judicial credibility construction"* (China's State Council 2014). Other official documents point to three main goals for the SCS: creating a "culture of integrity", solving economic problems, and improving governance (Ohlberg, Ahmed, and Lang 2017, 6).

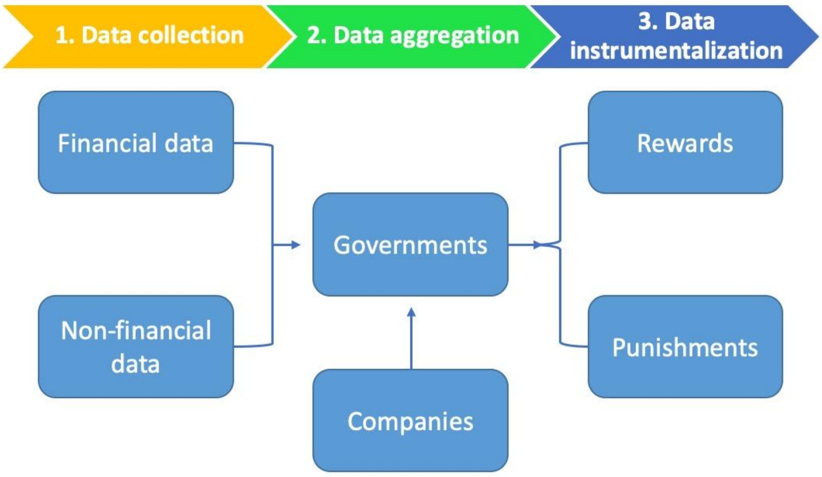
Table 1 - SCS' goals

Culture of integrity'	Economic problems	Governance
Restoring social trust and honesty	Boosting market efficiency and economic growth	Increasing government credibility
Rewarding good behavior, punishing the untrustworthy	Anti-counterfeiting	Improving information exchange within bureaucracy
Citizens' moral education (based on integrity as a traditional Chinese value)	Strengthening food and drug safety	Protecting private data (by regulating authorized access)
Educated online behavior ('Internet integrity')	Consumer protection	Fighting corruption

Source: adapted from Ohlberg, Ahmed, and Lang (2017, 6)'s figure 2.

This is an enormous task, just like the big data architecture behind it. To understand it, it is useful to examine how the system should look like once it is fully operational, and integrated. To implement the goals in Table 1, Liang et al. (2018) suggest that the SCS will execute three steps data-related processes that could explain allegations that it is an all-knowing "Orwellian" credit score (Horsley 2018).

Figure 1. SCS' data structure



Source: adapted from Liang et al. (2018, 21)'s figure 1.

As illustrated in Figure 1, First it collects financial and non-financial data on citizens, private, and public entities. Since it also collects data from companies like Alibaba, this can be everything from speeding tickets to online shopping behavior (Foreign Policy 2018). Second, the government will aggregate everything in a centralized master database (Schaefer and Yin 2019, 9–12). Third, it will instrumentalize this database to deploy a series of punishment and rewards mechanisms.

There are two main types of punishment and rewards, the point-based systems, and the blacklists

(Schaefer and Yin 2019, 12–16). This is usually what is referred to when talking about the SCS. However, to the system is still extremely decentralized and all its aspects, in fact, (Liu 2019, 23) calls it "Multiple Social Credit System", and there is virtually no information sharing among the systems, with the notable exception of the Court Defaulter Blacklist (Slater and Fenner 2011, 18).

According to Liu (2019, 26), 21 counties have so far implemented the point-based system. It can reward and punish citizens according to almost all aspects of life. Rongchen is a great example to illustrate this. This was the first county to adopt such a based system, and it is considered a model by the central government in this aspect to be followed. There, every adult resident starts with 1,000 points, and can be ranked from AAA to D on their scores, and according to their good or bad behavior city residents gain or lose points. There, they gain points for voluntary work and donating blood, for example, and lose points for breaking traffic rules and evading taxes (Gan 2019, 4–6).

The blacklisting system contains hundreds of different blacklists controlled by different state agencies, such as the Ministry of Ecology and Environment or the Tax Bureau, that may blacklist individuals and companies falling under their jurisdiction (Schaefer and Yin 2019, 12–16). However, the Court Defaulter Blacklist (CDB) is the only real enforcement tool that is available uniformly and national-wide (Gan 2019, 8–9).

The CDB was created to address the enforcement of court judgments. Predominantly, people and institutions are included in this blacklist for not repaying a debt, even after the court determines that they do have the financial means to do so (Liu 2019, 23). Courts in all administrative levels have the power to do so, and this can have very harsh consequences going much beyond the typical restriction to credit (Dai 2018, 33).

First, because it integrates with the other blacklists and municipal systems, it means the person will receive punishments across the board. The person can't buy high-speed train and plane tickets, they can be sometimes barred from job promotions, and their children might be blocked from attending private schools (Planet Money NPR 2018a). In an attempt to shame them, some counties even display the list with the people's faces in large and public billboards outdoors (Planet Money NPR 2018b). Unfortunately, there is no specific literature available about the CDB's nuances, but since this will be central for the methodological framework, more information about the CDB will be brought.

2.3. Autocratic stability and the Social Credit System

This part builds on the existing literature on autocratic stability to point how the SCS seems tailored to strengthen *regime stabilization, subsequently expanding the SCS co-optation, legitimation, repression dynamics individually.*

Slater and Fenner (2011, 14) argue that achieving stability is a government's task that goes beyond simply overcoming crises, but mostly avoiding them, or at least resolving them in the regime's favor. As discussed by the mainstream literature on autocracies, the regime stabilization is a process that even democracies go through. However, there have been many models trying to explain the fundamental differences of the consolidation processes between autocracies and democracies (Schedler 1998; Göbel 2011; Davenport 2007a, 2007b; Goldstein 1978).

In this sense, Gerschewski conceptualized one of the first models regarding the stabilization processes in autocratic regimes, the so-called "three pillars of stability". According to him, this framework is meant to enhance the regime's stability and survival, and is composed of three static pillars: legitimation, co-optation, and repression, which interact within themselves and with the others in a dynamic fashion. Given its integrative and dynamic approach to regime stabilization, I argue that the SCS fits the Gerschewski (2013)'s three pillars by the letter, hence its importance to understanding the SCS' objectives and implications.

Figure 2 below summarizes Gerschewski (2013)'s static definitions from each one of them (Dukalskis and Gerschewski 2018, 13). He defines co-optation as the regimes' ability to hook the elites to itself by material inducements, rewards, and policy concessions. Nevertheless, in addition to the author's original idea, I argue that the co-optation pillar ties not only elites but also the rest of the population by cultivating citizens' dependence on the regime (Slater and Fenner 2011, 22), as a sort of generalized form of clientelism. The legitimation pillar is activated by boosting people to support the government by showcasing the regime triumphs or by reaping an ideology that substantiates the autocrats' claim to power. Finally, the repression pillar consists of deterring activities that the state finds threatening.

Figure 2. Co-optation, Legitimation, and Repression

Source: Self-drafted based on Gerschewski (2013), Dukalskis and Gerschewski (2018) Slater and Fenner (2011).

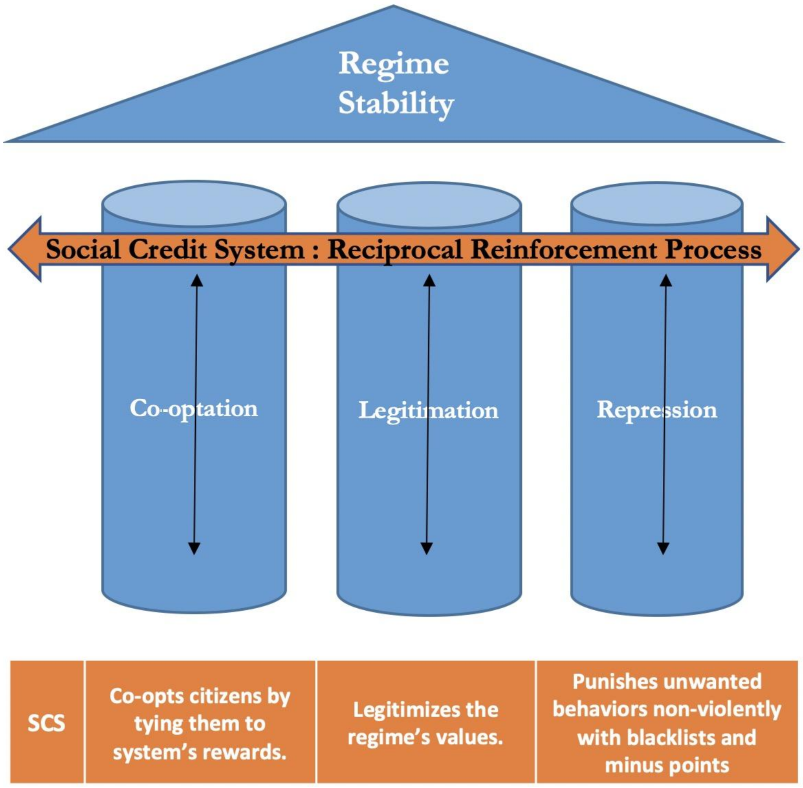
Furthermore, other than those static three definitions of each pillar, Gerschewski (2013, 24-30) expands on the interaction between and within those pillars to decipher the composition of the process of autocratic stabilization. Hence, he articulates the existence of three processes: exogenous reinforcement (among different pillars), endogenous self-reinforcement (within the same pillar), and reciprocal reinforcement (among different pillars). In other words, the autocrat's policies and actions do not arise in a vacuum, instead they interact in different forms.

On the one hand, the author considers that Legitimacy and Co-optation tend to be endogenous self-reinforcing processes, this is because investing is more sustainable, and their costs reduce with time. On the other hand, some legitimation policies often go beyond this, Gerschewski (2013, 25) highlights that it also has the power to reinforce the other pillar. For example, enhancing the population's satisfaction by delivering better public services sustainable reinforces legitimacy itself, but it also cuts the costs to co-opt the elites, appease potential opposition and reduces the need to repress. According to the author, processes like those are ideal to sustain stability.

Lastly, Gerschewski (2013, 28) paints a more intricate picture with a complex dynamic between repression and legitimation that often leads to unintended consequences. The author asserts that repression is usually not sustainable and represents an exogenous reinforcement process, that is when a regime represses its population, it is indeed mitigating the risk of insurgency and reducing the costs for future co-optation, but it is also simultaneously spending its legitimacy.

Interestingly, the SCS seems to incorporate Gerschewski (2013)'s theoretical framework features neatly. As illustrated in Figure 3, when fully integrated, the system can trigger sustainable reciprocal reinforcement processes to legitimize the regime's values, co-opt citizens by tying them to their rewards. Most importantly, it punishes unwanted behaviors from low scorers and blacklisted individuals in a very elegant way because it does not harm them physically. Instead, the SCS deprives citizens of basic rights, and from would-be advantages. Additionally, it is also very likely that the low scorers would blame themselves for their low scores once the SCS' "integrity values" (Ohlberg, Ahmed, and Lang 2017, 6) are sufficiently internalized in society. In fact, according to Dai (2018), this logic would be particularly efficient in China for its firm reputational-based society.

Figure 3. The three pillars of stability and the SCS



Source: Self-drafted and adapted from Gerschewski (2013, 23).

All in all, the SCS could help strengthen autocratic stability in multiple ways and has great potential to deepen the Chinese Communist Party's grip on power, hence the importance to understand it holistically. Next, this chapter will demonstrate theoretically how the SCS entrenches each pillar individually. However, my work will focus on the SCS' repressive aspect for two reasons because it would not be possible to tackle all those aspects empirically in this thesis. The reasons for this choice are twofold. First, because it is the most salient issue both in the media and from academia. Second, because it will center around the only completed integrated component of the system, the CDB, with clear repressive features.

2.3.1. SCS and co-optation

Can the SCS be considered a co-optation tool?

As previously discussed, the co-optation pillar consists of offering different advantages to citizens and elites, but only to hold power over them. This power appears when the regime threatens to withdraw those advantages when needed as an effort to keep those groups in line with the government's interests.

This autocratic form of clientelism completely aligns with the SCS' reward system. For example, in Rongcheng's municipal point system, plus points enable citizens to access exclusive free public and private services, better credit facilities, cheaper public transport, shorter waiting times for hospital services, etc. (Liang et al. 2018; Foreign Policy 2018; ABC News 2018). Other than material benefits, citizens can also enjoy the status of simply having a higher point score than their friends. The point is, that once citizens start profiting from any of those rewards, they will be much more likely to defend it, and its values, because they would be already strategically bonded to it. In this logic, high scorers will maintain their high scores by supporting the CCP values.

2.3.2. SCS and legitimation

Can the SCS be considered a legitimation tool?

Under the SCS' context, one can think about the many ways in which the SCS does this. To stay within the rewards logic, high scorers will maintain their high scores inasmuch as they support the values the CCP promotes throughout the SCS. Furthermore, China's SCS pledges concern for performance and transparency by increasing citizen engagement, given that even state structures, especially at the local level, are also subject to the system. This is an important aspect of the system for strengthening the regime's legitimacy. This would fit the idea of regimes fostering "passivity and political indifference among most of the population", claiming to have well-performing or successful economies, so people would conform to it. This type of regime adopted the performance mechanism (Dukalskis and Gerschewski 2018, 8-10).

The SCS could also be used for enhancing responsiveness, and representativity, at least at a local level. This is because the system punishes local administrators, on a personal level, for mismanagement. 2017 media reports pointed to more than 100 city, county, and country governments included in the CDB (Hongri Zhang 2017), and that more than 170,000 were blocked from senior management positions countrywide (Supreme People's Court of China 2017). The regime has even encouraged narrowly targeted protests to identify social grievances, to monitor lower levels of government, and to remedy the weakness of its political system (Chen 2012; Dimitrov 2008; Lorentzen 2013; Li 2019, 4). In fact, Meng et al. (2017), and Chen et al. (2016) elaborated large-N field experiments in the local level that hinted that the use of ICT policies as a source of authoritarian responsiveness is already reality in China. The first found that provincial and prefecture-level leaders are very likely incorporate formal offline and informal online citizens' suggestions into policy. The second suggests that most county administrations are very responsive to citizens' demands made online, particularly when they threaten Collective Actions. Henceforth, both studies show that the CDB fits this trend accordingly.

All in all, this would match the democratic-procedural mechanism, legitimation would occur through institutions of democratic nature, such as the holding of elections, existence of different parties, parliaments, and courts, only that in this case they would be manipulated or used as a tool for co-optation and repression (2018, 10-16). To this Chen and Cheung (2017) also argue that ICTs such as the SCS may empower citizens to challenge state authority and enhance state responsiveness to citizens' demands that together result in significant gains for the regime's legitimacy.

2.3.3. SCS and repression

Can the SCS be considered a softer form of repression?

The SCS repressive pillar is the focus of this work, hence this segment will build on the autocracy literature to address how the CDB can be considered a softer form of repression. In general, repression can be generally understood as "some form of coercive sociopolitical control used by political authorities against those within their territorial jurisdiction".[3] Goldstein (1978) and Davenport (2007a, 2) provide two different methods to define repression. The first is when the regime restrains the citizens' civil liberties by executing arrests or limiting freedom of expression, association, and belief; whereas the second type targets the individual's life and integrity, like torture (Davenport 2007a, 2; 2007b, 487).

In terms of intensity, it is difficult to draw a line on where a regime can be considered as a high or low repressive one. Johnston (2012), for example, mentions seven characteristics of High Capacity Autocracies, one of them is as highly developed social control, particularly when referring to China. However, a more suitable distinction for this thesis is indicated by Gerschewski (2013, 21), and originally suggested by Levitsky and Way (2002). They specifically separate between high and low-intensity repression according to the target and the form of the violence that has been imposed and suggest measurement from different databases (Gerschewski 2013, 21; Levitsky and Way 2002). High repression regards violation of an individual's physical integrity, whereas soft repression translates into less visible forms of coercion, such as "surveillance, censorship, harassment of journalists and activists, and the use of administrative procedures to prevent opposition gatherings", elements present in many societies (Dukalskis and Gerschewski 2018, 13).

This low-intensive repression logic might be the clearest way to which the SCS fits into the three pillars framework. The CDB in particular has parallels in the literature for autocratic censorship and contentious politics, where researchers analyzed how autocrats have recently used ICTs to repress the population. Hassanpour (2014) pointed to Mubarak's use of media disruption to mitigate revolutionary unrest investigated during the Arab Spring in Egypt, while Gohdes (2015)

[3] Goldstein 1978 as cited in Davenport (2007a, 2)

uses data from the Syrian civil war to investigate the correlation between increased military activity, and internet shutdowns as responses to periods with a higher level insurgency. Their results point to a clear instrumentalization of internet shutdowns as a softer and complementary form of coercion against the opposition.

Other comparisons can be drawn from King, Pan, and Roberts' (2013) results in the censorship of undesirable online content in China. They assert that the CCP tends to delete protest-related posts, but allows posts with specific types of criticism[4]. On a similar note, Qin, Strömberg, and Wu (2017) gathered over 13 billion posts to suggest that social media can be used as a surveillance tool to predict when protests would be happening and that the CCP uses this to counterweight menaces to the stability of the regime. All in all, those examples help situate the SCS' CDB within the broader literature addressing new ICT-related types of low-intensity autocratic repression.

2.4. Contentious politics, blacklists, and Collective Actions

Can the SCS' Court Defaulter Blacklist be considered a response to Violent Collective Actions as a form of repression?

As referenced in basic facts, being blacklisted already seems like a harsh punishment in itself. However, to call it repression, one would need to verify that it does act as a form of non-violent coercion against dissidence, as an extra way for the Chinese Communist Party (CCP) to bend the population over its will. Instead of arresting, or killing, one might be blacklisted as a new, and lighter way to be punished for engaging in Collective Action movements for example. This logic will be central to the upcoming methodological framework, but first, one needs to answer, if CDB could indeed be considered a response to Violent Collective Actions.

To define Collective Action, Zhang, and Pan (2019, 8) follow McAdam et al. (2003, 5)'s definition. For them, a Collective Action is an episodic event (not a regular meeting) with at least three people physically present, targeting political or economic power-holders, making a contentious and public claim that affects the interests of at least one of the other three. From now on, the term Collective Action will be preferred over protest.

The contentious politics literature has a lot to say about the nuances of the frequent repressive reactions from non-democratic regimes' towards protesters, particularly that Collective Actions very often trigger a reaction from the government (Goldstone and Tilly 2001; Tarrow and Tilly 2007; Gurr 1986). This is because protests have the potential to jeopardize regime stability, and

[4] According to King, Pan, and Roberts' (2013, 3), there is a tendency to allow posts criticizing local government corruption, and problems regarding service delivering.

when large enough, even regime survival. In many autocratic regimes, more often than not, this threat is met with violent repression.

However, violence might not always be the most suitable tool available for regimes to repress such incidents. Depending on the protest nature and scope, they also might be a sign of decreasing legitimation; hence violently repressing them might harm legitimacy even more in the longer run (Levitsky and Way 2010, 58; Gerschewski 2013, 21). Thus, the regime might try different ways to repress these protests. My argument is that one of those days could be blacklisting individuals.

According to Liu (2019, 3), this openness to employ a less violent way for repression has been growing among autocracies; instead, some of them tend to try to avert them deliberately (Brumberg 2002). Liu (2019, 3) points to China's "multifaceted nature of contentious politics" to highlight the diversity of repressive state responses in the country. He asserts that this happens because of the multitude of actors that might be involved, both in and out of the CCP.

Here there is a powerful, diverse, and decentralized autocratic state structure that is willing to innovate, and often eager to repress. In this context, given that anyone that does not comply with court orders can be blacklisted, one could easily assume they are being also used as a low-intensity alternative form of coercion. This could happen, for example, before the use of higher intensity forms of repression, like arrests, or killings. Particularly because inflicting the latter could bear the aforementioned legitimacy costs.

Last, it is also important to note that this assertion (blacklist = repression) can only be plausible once assumed that it does not extrapolate other determinant factors inherent to the CDB. Take its debt repayment aspect for example, this assertion would hold regardless of the number of debt people owe if verified that blacklists are disproportionately used in more rebellious regions. This concern will be addressed methodologically in the next chapter.

To sum up, there are three main theoretical take-ways from this chapter. First, once fully operational, the SCS has, in theory, the potential to strengthen the stability of the Chinese autocratic regime by simultaneously legitimizing its authority and values, co-opting citizens with rewards, and repressing unwanted behavior, as well as other types of dissidence. Second, the SCS's Court Defaulter Blacklist can be considered a new form of repression. Third, Collective Action can trigger repression, hence Collective Actions may also trigger blacklists. Altogether, they will serve as the basis for the subsequent methodological framework.

Chapter 3: Methodological framework

The theoretical discussion presented in the last chapter framed the lack of specific literature to address how the SCS' would function under an autocracy lens. To fill this gap, on a theoretical level, I addressed how the SCS could affect autocratic stability as a whole, focusing on how the SCS's Court Defaulter Blacklist can be considered a state reaction to Collective Actions, as a non-violent form of repression. Next, building upon those theoretical assumptions, the subsequent part will test the following research question empirically:

Research question: Do more Violent Collective Actions lead to more Court Defaulter Blacklists?

More specifically, I will test if the number of Violent Collective Actions (VCAs) positively affects the number of people placed on the Court Defaulter Blacklist, thus the main hypothesis will be:

H1: In Chinese counties with similar features, a higher number of Violent Collective Actions (X) is associated with a higher number of people placed on the court defaulters blacklist (Y).

This effect's existence would be the first empirical evidence to back up claims that the SCS can also be a tool for repression and to serve as a plausibility probe for potential large-N analysis. Additionally, if H1 holds this would cast doubt on the official Chinese Communist Party's rhetoric that the system is in place only to enhance societal trust and financial compliance.

3.1. Research Design

Throughout the next pages, this chapter will present the methodological framework in two phases. The first phase outlines the general logic of the chosen method, and the second applies this logic to this thesis' research design.

3.1.1. The co-variational analysis method

To test if H1 holds, I will deploy a case study based on a cross-sectional co-variational analysis (COV) as outlined by Blatter and Haverland (2012, 33–78). The aim here is to build cumulative and iterative empirical research profiting from both quantitative and qualitative approaches. This design aims at finding out the effect of an independent variable (X) on a dependent variable (Y) across cases happening in the same time interval given that the relevant control variables are held constant.

This approach is relevant for building theoretically oriented studies, and for developing applied research around newly introduced policies where data is often very scarce, as for the SCS. To this end, the COV analysis is perfect to get the most out of the large datasets available for

Collective Actions, and the CVs to get the most out of few CDB observations. This is a central reason to choose this small-N design instead of a large-N is connected to the serious difficulties to gather data for the CDB (DV).

Additionally, according to Gerring (2006, 152–172) case study based on co-variational analysis also harnesses other large-N experiment's strengths to understand the effects between two variables in social phenomena. In both large and small-N analysis, the relationship between X (treatment) and Y (outcome) can be established, only if the other factors influencing X are properly controlled for.

To illustrate this, let's imagine that a researcher wants to become the dictator of an imaginary country. She wants to test if more female leaders in a county lead to more female university graduates. She randomly nominates female mayors for half of the counties in her country (treatment group), while doing nothing to the other half of the counties (control group). After some years the female-led counties show a significantly larger amount of graduated females. However, before establishing a relationship between her policy and the number of female graduates, the researcher-dictator considered what else might have naturally influenced the number of female graduates independently of her policy. Since previous studies indicated that income and pregnancy rates influence the number of women graduating from university, she compared only counties with similar incomes and pregnancy rates. She is a dictator but still knows that different conditions might spur the relation between her policy and the number of graduated women.

While COV analysis follows a similar logic to the imaginary study outlined above, it also has differences. Blatter and Haverland (2012, 38) point to a central divergence between them when it comes to the choice of the counties or the case selection. Within experiments, researchers can hand-pick the right cases in such trials, whereas purely observable social studies do not have such an advantage. This limits the ability to manipulate both the treatments and the controls to the cases observable in society only.

For this reason, Blatter, and Haverland (2012, 41) point out that the case selection strategy is arguably the most crucial of COV case study analysis. As it will be seen in the next pages, selecting the right cases is central to validate the relationship between X and Y and will represent most of the operationalization of my work. For the authors there are two basic criteria for selecting the cases properly:

The first one is picking cases that vary as much as possible according to the treatment X (ex: female-led counties vs. male-led counties). Blatter and Haverland (2012, 44) indicated three modes of comparison, spatial/cross-sectional (county vs. county like the one example above), intertemporal (before vs. after intervention), and a combination of both.

The second basic criterion is to match only similar cases according to the confounding/control variables (ex: poor counties with high pregnancy rates). Within social sciences, according to Gerring (2006, 131), this type of case selection strategy works, to a limited extent, as a randomized experiment, since there is no intention to measure everything affecting the relation between X and Y. Instead, it aims at neutralizing unidentified factors across the treatment and control groups by randomization.[5]

Regarding the second step, Gerring (2006, 133) notes that the threshold defining who belongs to each category needs to be carefully considered. Following the example above, the question would be how to define a threshold to consider a county poor. Since social categories are not always black and white, the intention is to be as deep as necessary but remain as wide as possible regarding the categorization of each case. According to Gerring, this is a common trade-off in many studies, and it usually is not harmful as long as the hypothesis to be tested is always kept in mind.

3.1.2. Applied method

Applying this logic to my analysis, the independent variable will compare the number of VCAs within 6 counties (X). The dependent variable relates to repression, proxied by the number of people placed under the SPC blacklists (Y) within these same 6 localities. Once we keep other possible confounding factors invariable within those counties, divergences in the number of CDBs (Y) are hardly explained by other determinants.

As illustrated in Table 2 below, under those circumstances, there will be supporting evidence for the H1 if the number of blacklists is decisively higher in those counties with higher VCA. However, even if the results appear as outlined on the table, it is important to note that this relation between VCA and the blacklists does not invalidate other factors to influence the blacklists either. This would mean that Violent Collect Actions influence the CDB only under specific conditions, namely keeping the characteristics concerning the control variables unchanged among all selected cases.

[5] When referring to typical exact matching design, Gerring (2006, 135) suggests that "in a situation in which the set of matching variables includes some, but not all, confounders, matching may produce better causal inferences than regression models because cases that match on a set of explicitly selected variables are also more likely to be similar on unmeasured confounder". Even if those comparisons are valid in spirit, it is important to highlight that one should be careful to draw them among quantitative and qualitative methods since they can have, sometimes, different functions and objectives.

Table 2 - COV analysis methodology

Case	Control Variable 1	Control Variable 2	Control Variable 3	Independent Variable: Collective Action	Dependente Variable: Blacklisted individuals	
1	SAME	SAME	SAME	HIGH	HIGH	SUPPORTING EVIDENCE
2	SAME	SAME	SAME	HIGH	HIGH	
3	SAME	SAME	SAME	HIGH	HIGH	
4	SAME	SAME	SAME	LOW	LOW	
5	SAME	SAME	SAME	LOW	LOW	
6	SAME	SAME	SAME	LOW	LOW	

Source: Self elaborated.

In one way, this shrinks the range for generalization, but in the other way, it also makes the assertion more plausible within the types of cases that are similar (Blatter and Haverland 2012, 40). Nonetheless, I believe this approach is suitable because it profits from the extensive work from Zhang and Pan about CAs to reach a more targeted analysis for the blacklists, particularly because there is not much aggregated data available for the SCS and the Blacklists.

Considering this difficulty, in case the results point to a transparent relation between both variables, this work's main objective is to serve as a plausibility probe (Eckstein 1975, 128; Blatter and Haverland 2012, 40) for future large-N studies tackling similar arguments about the SCS.

This chapter focused on detailing the research question, the hypothesis. Subsequently, it centered on how this thesis intends to deploy a case study based on a cross-sectional co-variational analysis (COV) to address them. Those insights will be part to explanations provided next.

3.2. Data

Building on the previous parts, this chapter will articulate the details regarding the data used to measure both the dependent and the independent variables. It will only describe the original data sources but also give an overview of how they were collected, and the challenges behind them.

3.2.1. The independent variable: Collective Action

As detailed in the last chapter, Collective Actions could potentially trigger softer repression in the form of CDBs. For this reason, a higher incidence of Collective Actions can help point us to counties that are more likely to be repressed. Hence, there are different datasets available to measure Collective Actions in China. For instance, Goebel (2017) and Dimitrov and Zhang

(2017) use the Wickedonna Dataset[6] and the China Labor Bulletin[7], while Qin, Strömberg, and Wu (2017) compiled their own database. However, I will use Zhang and Pan (2019)'s Collective Action from Social Media (CASM-China) dataset for my analysis for distinct reasons.

First, because it covers the latest and the longest period (January 2010 to June 2017) in comparison to the other databases, yet it also has a great sample size with 142,427 events (Zhang and Pan 2019, 4). Second, due to its user-friendliness and free availability. Last, because it employs a great deal of AI and human hand-curated techniques to increase reliability. It does so by verifying if social media posts talking about a CA do correspond to events really happening offline. It also builds and checks its performance against the Wickedonna, China Labor's Bulletin, and other datasets to set up their algorithms, and to test its reliability. For those reasons, I will use Zhang and Pan (2019) dataset to measure Collective Action.

In a nutshell, the CASM-China is robust and works by collecting words that are related to CA. Subsequently, it uses deep learning to classify them based on images and text data to identify posts about CA offline in two stages. First, it distinguishes posts regarding complaints in general and, second, it identifies posts about CA. Additionally, it uses the respective posts to attribute location and time to CA uniquely. Moreover, following Almeida (2003), the CASM-China dataset classifies events into "conventional", "disruptive" and "violent". Conventional CAs regard events like strikes, public gatherings, and demonstrations accounting for 39%, while disruptive ones make up for 37%, representing more radical things like the occupation of land and buildings, barricades, and the deliberate interruption of electricity.

Lastly, the researchers coded actions like armed attacks and physical conflicts with government officials as "violent" Collective Actions (VCA). These events are coded as such when posts associated with them comprise preselected words in that categorization. VCA account for 24% and are coded this way when carrying any of the words in such a category, but also if they have markers belonging to the two other categories. Lastly, when possible, the CASM-China also individualizes the reasons behind each CA using the wording in the posts. It singled out 11 reasons varying from ethnic to fraud and even environmental reasons (Han Zhang and Pan 2019, 36–38). Unfortunately, individualizing any particular reason for the analysis would excessively shorten the number of cases, and turn the analysis no longer viable.

---

[6]This is a hand-curated Dataset created by two activists called Lu Yuyu and Li Tingyu. This is considered to be one of the greatest sources in the world for VCA in China. https://clb.org.hk/content/lu-yuyu-and-li-tingyu-activists-who-put-non-news-news.

[7] The China Labor Bulletin is an Hong Kong-based NGO that aims to help labor workers bargain with employers and advocate for their rights. One of their projects is to catalog labor VCA in China. https://maps.clb.org.hk/?i18n_language=en_US&map=1&startDate=2019-11&endDate=2020-05

3.2.2. The dependent variable: the SPC defaulter's blacklist

As outlined before, the SPC's blacklist is the only component of the SCS that has uniformed ramifications for all the other parts of the system countrywide. This is not the case for the municipal point systems or the other specialized blacklist. For this reason, the SPC's brings about the best opportunity to draw more consistent cross-sectional comparisons. Particularly because, since we are dealing with the same system, potential co-variations could not be attributed to differences within the blacklists themselves. As outlined before, unfortunately, this data is extremely difficult to get, and there is no dataset available. For this reason, I independently searched and coded it according to my research design's need.

There is a strong tendency to not post aggregate information, that is, how many people, their ages, or reasons to be blacklisted. Instead, "typical cases" with the person's name and specific wrongdoings are abundant. The focus is always on the "Lao Lai" individually[8]. This might happen by design to avoid questions like the one proposed in this thesis. The Chinese court system follows the general national administrative division: county, prefecture, provincial, and national level courts. Data on the national level is only available in official reports first published in mid-2018, and released monthly and annually since then. Unfortunately, those reports never disaggregate by provinces, prefectures, or counties, and their period is too short for a time-series comparison.

Furthermore, the national website of the Supreme People's Court publishes the total number of blacklists in real-time[9]. The names and the reasons for the blacklists are available on an individual basis[10], but no aggregate information is given. Unfortunately, the website also effectively blocked my attempts to automate the collection of this information via specifically programmed APIs. This selective transparency makes even more sense if we consider the SCS possible repressive intent.

Fortunately, aggregated county-level data is less difficult to find compared to the other administrative levels. At the lowest judicial level, county courts seem to be the primary blacklister to place and advertise people on their local partition of the blacklists system. When published on the other levels, it usually refers to the original county court entry. However, even on the county level, it is very difficult to find the aggregate numbers. The most consistent place to gather this information seemed to be the annual performance reports of individual county

[8] Pejorative nickname given to the people placed on the court defaulter's list.

[9] http://zxgk.court.gov.cn/

[10] This means that it is possible to search only once you have someone's name and social security number, one by one, never aggregately.

courts. It outlines other court statistics like the total number of opened and closed cases, people sentenced, arrested, and sometimes, it also indicated the total number of people placed on the Court Defaulter's Blacklists. For those reasons, I will use county-level court data to address my research question.

However, mining this information is by no means a straightforward process. Those reports are not always available, and more often than not, when available, they do not provide information on the blacklists. After many different attempts to get this information, I found that the most effective way to find them was by associating the specific keywords (Table 2), inserted either on the county court website or on the county administration websites.

In all the cases, I needed to add the "county name + court" followed by one of the terms in Table 3. In few occasions, there has been a batch list with name, and sometimes photos of each person, these were counted manually[11]. Most counties stopped posting this information on their local annual performance report between 2017 and 2018, this was the exact period that the national website[12] and national reporting started. For this reason, my analysis considers SPC's blacklists from reports containing information from the introduction of the policy from 2015 until 2017.

Table 3. Keywords for the Court Defaulter Blacklist's Search

Solving enforcement difficulties[13]	基本解决执行难
Court Defaulter Blacklist	失信被执行人
Blacklist	黑名单
Work report	人民法院工作报告
Laolai	老赖
List Laolai	老赖清

[11] List with the links to county courts reports here: https://bit.ly/3l5uO1c
[12]See note 9.
[13] This is the official name of the policy document that created the court defaulters blacklist.

Lao Lai Exposure Station	老赖曝光台

Source: Self elaborated.

The data used here is reliable because it comes exclusively from the official website from either the county courts or the county administrations. Beyond CA and the SPC defaulter's blacklist, other datasets will be used to incorporate relevant control variables, namely: GDP per capita, urbanization, population, and internet penetration. They will be addressed individually throughout the next pages.

3.3. Operationalization: case selection and control variables

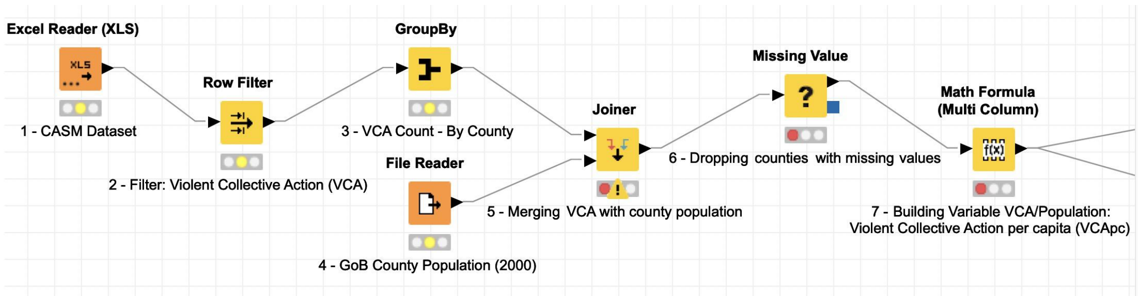
To execute the proposed cross-sectional COV-analysis' logic, this part will operationalize Gerring (2006, 131–34)'s cross-case technique in three phases. First, it will expand on how and why the variable Violent Collective Action per capita (VCApc) was put together. Second, it will do the same regarding each control variable used. Finally, it will end by applying all the variables' thresholds to match only the cases that fit all the necessary criteria to be in the final case selection.

As stressed in the research design chapter, the relationship between VCApc (IV) and the SPC defaulter's blacklist (DV) concentrates heavily on the correct operationalization of the case selection under Gerring (2006, 131)'s quasi-randomized experiment logic. This happens because the independent variable is fundamentally based on the features of the cases to be selected. To achieve this, I will use Knime (Berthold et al. 2009), a user-friendly information miner software that works like R and Stata. Furthermore, the case selection will be built upon the theoretical discussions brought here and will combine them with Gerring (2006, 131–34)'s cross-case technique, a case study variation of a typical matching strategy. Its adaptation to my analysis will follow the six steps below:

1. ***Rank the counties based on VCApc;***
2. ***Assign treatment vs. control groups (highest VCApc vs. lowest VCApc)***
3. ***Identify relevant control variables;***
4. ***Dichotomize the all variable's scores (ex: large/small, high/low);***
5. ***Match only counties sufficing all IV and CVs scores/criteria;***
6. ***Compare blacklists between the high (treatment) and low (control) VCApc groups (case study analysis);***

3.3.1. Violent Collective Action per capita (VCApc)

Figure 4. Counties ranking based on VCApc



Source: Self-drafted with Knime 4.1.2.

The steps to rank counties based on VCApc (Step 1) are outlined in Figure 4 above. Only the CA coded as "violent" to assure that only the most extreme cases would be accounted for (Item 2 in Figure 4). At this point, the sample has 14524 VCA distributed throughout 1162 counties and averaging 2,81 VCA per 100.000 inhabitants. Following the COV analysis logic, there are two main motivations to rank counties according to their number of VCApc, the first is to get the largest variation possible in the Independent Variable, and the second is to maximize time and resources looking for CDB data only from counties that would enhance the final case selection results' inference the most.

I believe this is important to increase the likelihood of finding events that would be more likely to be punished. This idea is in line with claims that violent claims are often prompted to be repressed in China (Selden and Perry 2010; Cai 2010). One might argue that if blacklists are less violent forms of repression, it would make more sense to also take less violent CAs. This would make sense as a large-N study because it would be able to capture the nuances of the relation between CAs and the Court Defaulter Blacklist more directly. However, in my analysis, there is not a large-N character. Here, it is important to highlight my intent to use the CASM-China dataset to point me as accurately as possible to the regions with the most rebellious populations for the final case selection. This way I can tailor the analysis to identify repression originated in response to CA.

The precision to which the CASM-China dataset can locate where a VCA has happened varies from county-level, to prefecture-level, to provincial-level. It follows the GuoBiao (GB) codes for the administrative divisions from the Chinese Academy Of Surveying And Mapping (1997)[14]. In this analysis, I grouped the VCA (VCA) by the number of times they happened in each of

[14] The GOB code is not the same as the postal code, instead, it identifies and standardizes all the administrative units in China, province, prefecture, and county. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/EOH3FV.

those administrative divisions, where I considered only the events that were successfully parsed to the county-levels (Item 3). The reasons for this were twofold. First, to get the values at the prefecture-level or the provincial-level, I would need to collapse too many events together. This would mean gains in terms of external validity but could jeopardize internal validity and the overall precision for the case selection and the analysis' final results. Second, because preliminary checks suggest that there is more information available about blacklists (Y) on the county level than on the other levels.

Berman (2017)'s data on the Chinese population disaggregated to the county level was merged with the data on VCA, as illustrated in Items 3, 4, and 5. To this point, it is clear that a city with a million inhabitants will have a higher absolute number of such events than a city with 500.000 inhabitants. To address this, the VCA of each county were divided by their respective population. The result would be the proportion of VCA per 100,000 inhabitants (VCA per capita).

The missing values obtained from the merge were due to changes in the name of counties or the absence of VCA's happening in the period of collection. In such cases, these counties were left out of the sample in Item 6 to obtain the single variable: VCA per population (VCApc), in Item 7.

Berman (2017)'s data from the 2000 census is considerably older than the most recent census in 2010, although it is easily available in terms of cost and user-friendliness in comparison with other sources.[15] Most importantly, it also uses the same GOB administrative code that allows for automated cross-combination of the VCA counts from each county and their respective population, as seen in Item 5.[16]

The variation between the population in the 2000 census and at the beginning of the Blacklist Policy in 2014 is not neglectable, but it should suffice to provide a baseline to build a suitable case selection. Additionally, because the 2010 census is occasionally available for consultation county by county, it will still be used in later stages of the analysis once the final selection is ready.

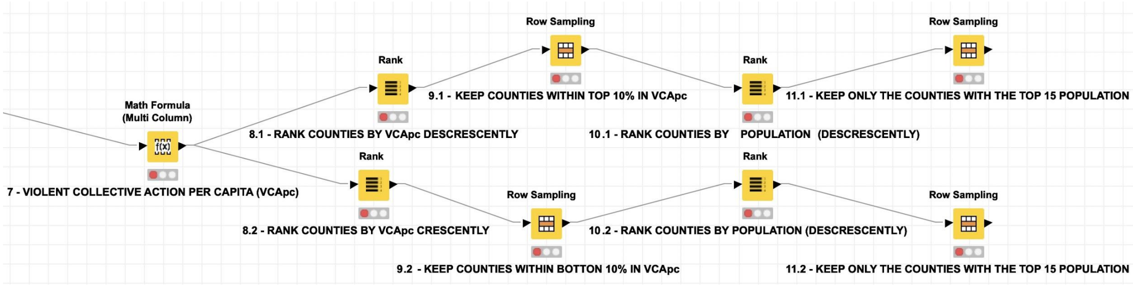
Lastly, based on the newly created VCApc, I followed ***step 2 by assigning cases to the highest VCApc group (treatment) and to the lowest VCApc group (control).*** To achieve this, Item 8.1 ranked the counties from higher to lower VACpc, and 8.2 ranked them lower to higher, as illustrated in Figure 5 below.

---

[15] http://www.stats.gov.cn/english/Statisticaldata/CensusData/rkpc2010/indexch.htm

[16] The CASM-China dataset locates VCA using the individual GOB administrative county code. If the population database does not contain the same code attached, it would not successfully connect the counties with their population.

Figure 5. Assigning counties to the treatment (High VCA counties) and control groups (Low VCA counties)



Source: Self-drafted with Knime 4.1.2.

3.3.2. Control variables

Gross Domestic Product per capita (GDPpc)

To follow up on step 3, I will start identifying the control variables. GDPpc considers the analysis needs to relate to the financial aspects of each county. This is important because, as explained before, the main reason for people to be placed on the CDB is not having paid their debts. The ideal way to control for this would be having an indicator for household debt, like one from the Institute of Social Science Survey 2017 or the CEIC,[17] but unfortunately, they do not provide data at the county level. For this reason, I will use GDPpc[18] as an instrument, because this data is available at the county level and it could, at least partially, capture this financial aspect.

Following the average national GDPpc in 2010, ***the threshold for assigning a county to high or low will be 30,808,000 RMB (step 4)***. The case selection will then focus on the lower-income counties because poverty is said to be often one of the reasons to instigate CAs, especially in rural China (Hurst 2004; Hess 2010; Ngai and Huilin 2010).

Urbanization

The reason to take urbanization as a control variable relates to the CASM-China dataset limitations to capture events happening in rural areas. This constraint exists because the CASM is primarily based on social media posts, this means that regions with less internet penetration would not be represented accordingly. Therefore it is reasonable to use urbanization as a control for this analysis, particularly because low levels of urbanization are commonly associated with low internet penetration levels in China (Wunnava and Leiter 2009; Dasgupta, Lall, and Wheeler 2005).

---

[17] https://www.ceicdata.com/en/indicator/china/household-debt--of-nominal-gdp

[18] Manual consultation by county using the CEIC data: https://www.ceicdata.com/en/china/gross-domestic-product-county-level-region

Nonetheless, Zhang and Pan (2019, 41) do underscore that their dataset performs relatively better than others when it comes to capturing CAs connected to rural land disputes. In fact, 23% of the events identified related to such contexts can be a manifestation of the increment in the use of social media observed in the last years in the country as signaled by McDonald (2016).

What's more, researchers have highlighted the importance of the CA in rural China on many occasions (Bernstein 2004; O'Brien and Deng 2015; Pu and Scanlan 2012). To capture those aspects, ***the analysis will dichotomize urbanization (step 4) by focusing on counties with equal or lower than 50,7% of urbanization rate (the national average) according to the China National Bureau of Statistics[19].***

County's population and VCApc

The reasoning for including the county's population as a control variable relates to a hint brought up in a first trial matching the VCA and the blacklists. At this stage, it is still not possible to infer trends, but it may hint that smaller cities have more blacklisted individuals on average. As shown in Tables 4 and 5, the Low VCA group (Table 4) has consistently larger population sizes, averaging 746 thousand inhabitants, almost double the High VAC Group' average.

Table 4 - High VCA group (Treatment)

High VCA Group								
Position	**County code**	**County Name (Pinyin)**	**Province**	**Population**	**VCAs**	**VCA/100.000 Inhab.**	**Blacklisted**	**Blacklisted/100.000 Inhab.**
1	622925	Hezheng Xian	Gansu	192.160	313	162,885	99	51,52
2	431230	Tongdao Dongzu Zizhixian	Hunan	215.296	122	56,666	142	65,96
3	430121	Changsha Xian	Hunan	745.238	317	42,537	756	101,44
4	542425	Anduo Xian	Xizang	32.443	13	40,070	119	366,80
5	433123	Fenghuang Xian	Hunan	365.694	121	33,088	187	51,14
6	140522	Yangcheng Xian	Shanxi	396.174	111	28,018	130	32,81
7	610331	Taibai Xian	Shanxi	51.142	11	21,509	110	215,09
8	513225	Jiuzhaigou Xian	Sichuan	56.167	12	21,365	26	46,29
9	210421	Fushun Xian	Liaoning	226.401	46	20,318	645	284,89
10	360121	Nanchang Xian	Jiangxi	986.031	186	18,864	5948	603,23
Mean				**326.675**	**125,2**	**44,532**	**816,2**	**181,92**

Source: Self elaborated

[19] Relative to the 2010 census and extracted from www.citypopulation.de.

Table 5 - Low VCA group (Control)

Low VCA Group								
Position	County code	County Name (Pinyin)	Province	Population	VCAs	VCA/100.000 Inhab.	Blacklisted	Blacklisted/100.000 Inhab.
1	530326	Huize Xian	Yunnan	872.361	1	0,115	2844	326,01
2	370826	Weishan Xian	Shangdong	677.893	1	0,148	1820	268,48
3	211321	Chaoyang Xian	Liaoning	614.057	1	0,163	1333	217,08
4	220322	Lishu Xian	Jilin	855.538	1	0,117	1602	187,25
5	511028	Longchang Xian	Sichuan	757.049	1	0,132	1296	171,19
6	411321	Nanzhao Xian	Henan	600.269	1	0,132	732	121,95
7	510322	Fushun Xian	Sichuan	1.208.556	2	0,165	1330	110,05
8	130828	Weichang Manzu Mengguzu Zizhixian	Hebei	509.169	1	0,196	452	88,77
9	220721	Qianguo'erluosi Mengguzu Zizhixian	Jilin	551.988	1	0,181	165	29,89
10	430922	Taojiang Xian	Hunan	819.381	1	0,122	195	23,80
Mean				**746.626**	**1,1**	**0,147**	**1176,9**	**154,45**

Source: Self elaborated

The reasons for this are unclear but it suggests that, at the very least, population size might have an effect on the number of people blacklisted and should be considered as a control variable. Additionally, comparing the group with less VCA averages 154 blacklisted people per 100,000 inhabitants (Table 4), while the group with more VCA averages 181 (Table 3). In line with H1, this is a first hint that cities with more VCA per capita may also have slightly more blacklists.

Next, I will use the aforementioned hint regarding population size ***to dichotomize the IV VCApc (step 4).*** Items 9.1 and 9.2 (Figure 5) keep only the counties with the top 10% the High VACpc and the counties with the bottom 10% Low VACpc groups. This is essential to obtain the largest variation possible in the independent variable.[20] After this point, each list retained only 116 counties each. The top and bottom 10% are good choices because they are far from having the same number of VCApc. In other words, even the least rebellious county in the treatment group (rank 116th) would still have much higher VCApc than the most rebellious of the control group (rank 1162th). This is much less risky than getting the top 50% and the bottom 50% for example.

To dichotomize the CV population (step 4), I first used items 10.1 and 10.2 to rank the top 116 counties, and the bottom 116 based on their population sizes. Subsequently, I used items 11.1 and 11.2 to keep two lists, one with the 15 counties[21] with the highest populations and the highest VCApc, and another one with the 15 highest populations and the lowest VCApc.

Internet penetration

Unfortunately, neither social media nor internet penetration data are available at the county level,

[20] As explained in the chapter about the data, I kept the most populated counties instead of the smallest ones because finding they tend to have more information on their blacklists available online.

[21] After different trials, 15 counties was the smallest number that allowed me to maintain enough cases to suffice the next phase of the selection with the remaining control variables.

but the latter is indeed available on the provincial-level[22] (Knoema 2019). This would be a good opportunity to test the CASM-China reliability concerning internet penetration but some attempts in this direction left too few cases to be analyzed and would not necessarily help to explore the dependent variable further. For this reason, the ***counties chosen are within provinces with higher internet penetration levels (step 4)***, except for Liling and Kaizhou. Yet, even those two are located in pocket areas with very high internet penetration. The first has almost 69% and the second 77,4%, this is much higher than the average by provinces of 45.3%[23].

3.3.3. Matching cases

To follow *step 5, I matched only counties sufficing all IV and CVs scores/criteria* by compiling the aforementioned information together in Table 5 and Table 6. The first aspect to note is that the high VCApc Group has a higher GDPpc on average with a combination of rural and urban counties. In the High VACpc Table 5 below, the counties in green came close to fulfilling all the criteria but did not belong to areas with higher internet penetration levels. The only counties that checked all the boxes to go to the final selection were the ones in orange.

Table 6 - High VCApc group

High VCApc Group									
County	VCAs	Pinyin Name	Province	2010 Population	VCA per 100.000 inhabitants	VCApc Ranking Position	GDPpc 2010 (RMB)	Ruralization (%)	Internet Penetration Rate 2010*
南昌县	186	Nanchang	Jiangxi	1,018,675	18.26	20th	31,367,000	57.9%	30.0%
苍南县	100	Cangnan	Zhejiang	1,184,643	8.44	59th	19,800,000	45.3%	60.5%
南安市	112	Nan'an	Fujian	1,418,451	7.90	69th	33,971,930	49.3%	62.6%
平度	98	Pingdu	Shangdong	1,357,424	7.22	73th	38,123,170	68.5%	43.9%
仙游县	**67**	**Xianyou**	**Fujian**	**824,707**	**8.12**	**83th**	16,906,220	63.2%	62.6%
双流区	58	Shuangliu	Sichuan	1,279,930	4.53	85th	43,940,000	37.9%	34.3%
许昌县	49	Xuchang	Henan	767,449	6.38	97th	22,458,540	72.9%	34.0%
阳新县	59	Yangxin	Hubei	827,631	7.13	99th	13,492,000	71.2%	42.5%
平阳县	**47**	**Pingyang**	**Zhejiang**	**761,664**	**6.17**	**105th**	23,421,000	50.7%	60.5%
如东县	65	Rudong	Jiangsu	995,983	6.53	106th	35,592,000	51.9%	51.0%
东台	68	Dongtai	Jiangsu	990,306	6.87	107th	36,616,000	50.5%	51.0%
福清市	67	Fuqing	Fujian	1,234,838	5.43	111th	38,651,370	61.9%	62.6%
慈溪市	57	Cixi	Zhejiang	1,462,383	3.90	112th	73,037,000	27.5%	60.5%
东海县	**60**	**Donghai**	**Jiangsu**	**952,668**	**6.30**	**116th**	20,696,000	58.1%	51.0%
岳阳县	42	Yueyang	Hunan	716,829	5,59	113th	19,809,000	65,8%	35.6%
Mean	**76**			**1,052,905**	**7.37**		**31,192,082**	**54.8%**	**49.5%**

Source: Self elaborated.

Second, within the Low GDPc Group, the majority of the counties are very rural and have a low GDPpc. Additionally, as explained, almost all the counties within this group are placed in provinces with lower internet penetration rates, whereas the opposite is true for the High VCA Group. Within the Low VCA group, all counties are within provinces with low internet

[22] Internet penetration rate can be generally understood as the percentage of the total population that uses the internet at any level (city, county, province, state, region, country) - https://knoema.com/CNIPS2017/china-internet-penetration-statistics-by-province

[23] The details, sources, and calculation for these numbers can be accessed here: https://drive.google.com/file/d/10VDbqGltYFK6QMUIS-qUmtZlwRZnADVX/view?usp=sharing

penetration levels, except for Liling, Kaizhou, and Xinghua (in blue). Cangshan and Ju Xian (in green) could be considered for the analysis because unfortunately, the number of blacklisted individuals for them was not found. The only counties that checked all the boxes to go to the final selection were the ones in blue.

Table 7 - Low VCApc group

Low VCApc Group									
County	VCAs	Pinyin Name	Province	2010 Population	VCA per 100.000 inhabitants	VCApc Ranking Position	GDPpc 2010 (RMB)	Ruralization (%)	Internet Penetration Rate 2010
兴化市	4	Xinghua	Jiangsu	1,253,548	0.32	1120th	30,025,000	54.1%	51.0%
开州区	6	Kaizhou	Chongqing	1,110,336	0.54	1052th	17,214,000	64.1%	43.1%
资中县	5	Zizhong	Sichuan	1,192,060	0.42	1071th	11,479,000	77.8%	34.3%
唐河县	4	Tanghe	Henan	1,282,262	0.31	1096th	15,208,990	71.1%	34.0%
云阳	3	Yunyang	Chongqing	912,912	0.33	1130th	11,983,000	67.8%	43.1%
怀远	3	Huaiyuan	Anhui	1,028,066	0.29	1129th	10,123,670	75.5%	33.9%
榆树市	1	Yushu	Jilin	1,160,568	0.09	1162th	21,765,000	75.8%	41.7%
富顺	2	Fushun	Sichuan	826,195	0.24	1153th	14,838,000	69.4%	34.3%
兰陵	4	Cangshan	Shangdong	1,161,932	0.34	1083th	15,175,890	68.6%	43.9%
莒县	2	Ju Xian	Shangdong	995,552	0.20	1148th	17,370,620	74.9%	43.9%
农安县	4	Nong'an	Jilin	960,759	0.42	1075th	19,919,000	81.3%	41.7%
醴陵	4	Liling	Hunan	947,387	0.42	1057th	27,737,000	52.6%	35.6%
忠县	2	Zhong	Chongqing	751,424	0.27	1142th	18,254,000	67.1%	43.1%
祁东县	3	Qidong	Hunan	979,855	0.31	1089th	14,138,000	67.4%	35.6%
莱阳	3	Laiyang	Shangdong	878,591	0.34	1088th	35,752,110	59.2%	43.9%
Mean	3			1,028,066	0.32		17,214,000	68.4%	40.2%

Source: Self-drafted.

This chapter concentrated on operationalizing the case selection strategy to fit the cross-sectional COV-analysis' logic. To achieve this, it iteratively started by ranking the counties based on the IV, followed by the assignment of treatment vs. control groups, it identified relevant control variables, subsequently dichotomizing the variable's scores, and it finalized by keeping only counties fulfilling all IV and CVs scores in the analysis.

Chapter 4: Case study analysis

4.1. Analysis and comparisons

This segment will critically compare blacklists between the high (treatment) and low (control) VCApc groups (step 6). It will start by describing the final case selection, and tie them back to the research question and the main hypothesis. This segment critically compares the relative validity of the evidence presented by the case study, elaborates on the extent to which the internal and external validity of the results can be drawn.

The final case selection consists of 3 cases belonging to the group with the High VCA group (Treatment) and 3 cases belonging to the Low VCA group (Control). All the counties have similar features for the control variables (Population, GDPpc, Urbanization, and Internet Penetration). The difference lies in the independent variable (VCApc), where the first group has high scores and the second has low scores. Following the COV analysis' logic, there is supporting evidence for H1, if the High VCApc counties (Treatment) also have significantly higher numbers of blacklisted people per 100.000 inhabitants than the counties within the Low VCApc Group (Control).

Table 8. below fills in the blanks for Table 2 to illustrate the results achieved here. It can be observed that the 3 counties selected on the High VCA Group of counties have an average of 528,9 blacklisted people per 100,000 inhabitants. This is more than 5 times lower within the Low VCA Group, where the average is 96 blacklisted individuals per 100,000 inhabitants.

Table 8 - Results from the co-variational analysis

<table>
<tr><th>County</th><th>CV1 - Population[24]</th><th>CV2 - GDPpc[25]</th><th>CV3 - Ruralization[26]</th><th>IV - VCApc ranking position[27]</th><th>IV - VCA/100.000 inhabitants[28]</th><th>DV - Blacklisted people/100.000 inhabitants[29]</th><th>Average blacklisted people/100.000 inhabitants[30]</th><th>Type of evidence</th></tr>
<tr><td>Pingyang
(Zhejiang)</td><td>761,664
LARGE</td><td>23,421,000
LOW</td><td>50,7%
RURAL</td><td>105th</td><td>6.17</td><td>470.68</td><td rowspan="3">528.94</td><td rowspan="6">Supporting evidence</td></tr>
<tr><td>Xianyo
(Fujian)</td><td>979.425
LARGE</td><td>16.906.220
LOW</td><td>63.2%
RURAL</td><td>83th</td><td>6.84</td><td>638.53</td></tr>
<tr><td>Donghai
(Jiangsu)</td><td>1.100.047
LARGE</td><td>20.696.000
LOW</td><td>58.1%
RURAL</td><td>116th</td><td>5.45</td><td>477.61</td></tr>
<tr><td>Kaizhou
(Chongqing)</td><td>1.470.757
LARGE</td><td>17.214.000
LOW</td><td>64.1%
RURAL</td><td>1052th</td><td>0.54</td><td>100.69</td><td rowspan="3">96.60</td></tr>
<tr><td>Liling
(Hunan)</td><td>1.011.279
LARGE</td><td>27.737.000
LOW</td><td>52,6%
RURAL</td><td>1057th</td><td>0.42</td><td>125.29</td></tr>
<tr><td>Xinghua
(Jiangsu)</td><td>1.545.838
LARGE</td><td>30.025.000
LOW</td><td>54.1%
RURAL</td><td>1120th</td><td>0.32</td><td>63.82</td></tr>
<tr><td colspan="9">Note: In 2010, the average GDPpc was 30,808,000 RMB, and the average urbanization was 50,7%.</td></tr>
</table>

Source: Self elaborated.

As it can be observed, the results do seem to provide supporting evidence to validate H1: In Chinese counties with similar features, a higher number of VCAs (X) is associated

[24] 2000 population census. The counties were selected from the top 15 with the largest population among both the high and the low VCApc groups.

[25] GDP per capita in 2010 (RMB). 2010 national average of 30.808.000 RMB.

[26] Urbanization in 2010. National average of 50,77% in 2010.

[27] Original position based on the 2000 population census and taken from 1,162 counties in the sample.

[28] Based on the 2010 population census.

[29] Based on the 2010 population census.

[30] The details and sources about the number of blacklists are available here: https://bit.ly/3l5uO1c.

with a higher the number of people placed on the Supreme People's Courts (SPC) defaulters' blacklists (Y). To paraphrase the English political economist John Stuart Mill:

*"If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or an indispensable part of the cause, of the phenomenon" (*Mill 1843, 455).

His logic, even if classic, can still be applied in modern contexts like the SCS. However, the cautionary note here is to avoid being categorical as to assume that the VCA single-handedly *caused* or represented a completely *indispensable part of the cause* for the higher number of blacklists.

4.2. Internal validity

To properly address this generalization, one needs to consider that a COV Analysis usually presents a trade-off: the deeper you analyze the variables, the fewer cases you can take on board. Hence, on the one hand, to address internal validity I chose to stick with the county-level analysis and not collapsed the data into prefecture or the provincial levels. In fact, I did not use events to which the CASM-China dataset was not able to parse the exact county where they happened. The aim was to achieve better levels of correctness according to the cases that were studied to avoid conceptual stretching (Sartori 1970; Lijphart 1975, 169).

On the other hand, it is clear that this choice also harms the correctness of the hypothesis concerning the population of inference (all the counties that were not studied). Therefore, it is crucial to acknowledge that the external validity of the results presented is limited. It rests upon the representativeness of the original CASM sample from where the case selection derives from and can be generalizable for cases that have features, namely poorer, rural and larger counties.

Furthermore, the SCS is a policy innovation and there is not much research about it, neither small nor large-N. For such cases, the use of plausibility probes studies like the one showcased here are better equipped to focus on internal validity (Blatter and Haverland 2012, 229; Gerring 2006, 217). Additionally, to minimize the "many variables, small N problem", I increased the number of cases as much as possible and tried to include counties from different provinces to have more geographical difference (Lijphart 1975, 163) and avoid spill-over effects (Lin, Chang, and Zhang 2015).

I also combined two variables, this is the case of Population and VCA and Population and GDP, this is useful to reduce the property space in the analysis (Barton 1955). Subsequently, I used the control variables to build a case selection of truly comparable counties. Finally, I focus the analysis on the key variables by choosing only the VCA because those would more likely correspond to the blacklists and hence, to repression.

4.3. External validity

To enhance external validity and the plausibility of the effect proposed, I probed smaller versions of the same COV analysis[31]. However, this time, I changed the values for the control variables to verify if the results would stay the same (Blatter and Haverland 2012, 230; George et al. 2005, 34–35). The case selection followed a similar logic from the main analysis because I kept the cases that fulfilled all the conditions in line with the control variables.

As seen in Table 9 below, in the first try, I searched for two richer counties while keeping all the other variables the same. Other than finding the number of blacklists, the challenge here was to find rich enough counties within the Low VCApc group because, as explained before, most of the ones in this group are less urban and less wealthy. Nonetheless, for the two compatible cases found, the evidence did align with the previous results. This could hint that, again, H1 might hold in counties with higher GDPpc within a large-N analysis.

Table 9. External validity check for COV analysis: Higher GDPpc

County	CV1 - Population	CV2 - GDPpc	CV3 - Urbanization	IV - VCApc ranking position	IV - VCA/100.000 inhabitants	Blacklisted people/100.000 inhabitants	Type of evidence
Nanchang (Jiangxi)	1,018,675 LARGE	31,367,000 HIGH	57.9% RURAL	20th	18.26	548.26	**Supporting evidence**
Weishan (Shandong)	633,357 LARGE	33,894,730 HIGH	61.5% RURAL	1156th	0.15	215.52	

Note: In 2010, the average GDPpc was 30,808,000, and the average urbanization was 50,7%. Average Population of the sample is 531.567 inhabitants and from the Low VCApc is 584,034

Source: Self elaborated.

Subsequently, to test for urbanization, one could take the same steps but this time select only urban counties, while keeping the population sizes larger and the GDPpc lower. Unfortunately, this combination was not found in the Low VCApc group. Maybe because larger urbanization levels are more often connected to higher GDPpc levels. Additionally, the number of blacklists was not found for the few poor and urban counties Low VCApc counties remaining.

As illustrated in Table 10 below, the closest I reached to this ideal combination had both higher GDPpc and higher urbanization level. This probe was the first one to show disconfirming

[31] All the tables and sources from this segment can be verified here: https://docs.google.com/spreadsheets/d/1Yjp9dp-VXPJowfsMEbyeoilmdSEXN0gxtVSeH1ZAEMA/edit?usp=sharing

evidence. Counterintuitively, the county with less VCApc had more Blacklists. However, as highlighted before, it is important to note this specific county, Rongchen, is widely recognized as a model city for the Social Credit System and is likely to be an outlier, particularly because this is one of the very few counties within 116 in the Low VCApc Group that is both rich and urban.

Table 10. External validity check for COV analysis: Higher GDPpc, and Higher Urbanization

County	CV1 - Population	CV2 - GDPpc	CV3 - Urbanization	IV - VCApc ranking position	IV - VCA/100.000 inhabitants	Blacklisted people/100.000 inhabitants	Type of evidence
Cixi (Zheijiang)	1,462,383 LARGE	73,037,000 HIGH	27.5% URBAN	112th	3.90	301.15	**Contrary evidence**
Rongchen (Shandong)	714,355 LARGE	95,103,450 HIGH	49.1%% URBAN	1103th	0.29	439.00	
Note: In 2010, the average GDPpc was 30,808,000, and the average urbanization was 50,7%. Average Population of the High VCApc sample is 531.567 inhabitants and from the Low VCApc is 584,034							

Source: Self elaborated.

At last, I drew 2 random samples of 8 counties using Knime (Berthold et al. 2009). One sample from within the 116 High VCApc counties and one sample from the 116 Low VCApc obtained after the items 9.1 and 9.2 from Figure 5. From within all those 16 counties, I skipped the ones I could not find the numbers of Blacklists and the ones that were already used before in my analysis. I stopped the manual search as soon as I found the first number of blacklisted individuals for each of the groups.

As shown in Table 11 below, the two counties are fairly different and fairly representative of the trend seen so far for each group. Kunshan, from the High VCApc Group, was richer and urban and Lufeng, from the Low VCApc group, was poor and rural. Again, following the logic of H1 again, the number of blacklisted individuals was much higher in the county with higher VCApc than in the one with lower VCApc. This is a further indication that H1 would hold in a large-N analysis.

Table 11. External validity check for COV analysis: Randomly assigned counties

County	CV1 - Population	CV2 - GDPpc	CV3 - Urbanization	IV - VCApc ranking position	IV - VCA/100.000 inhabitants	Blacklisted people/100.000 inhabitants	Type of evidence
Kunshan (Jiansu)	1,644,860 LARGE	104,413,000 HIGH	32% URBAN	37th	4.07	410.49	Supporting evidence
Lufeng (Yunnan)	422,770 SMALL	20,179,000 LOW	64,7% RURAL	1126th	0.29	79.48	
Note: In 2010, the average GDPpc was 30,808,000, and the average urbanization was 50,7%. Average Population of the sample is 531.567 inhabitants and from the Low VCApc is 584,034							

Source: Self elaborated.

Chapter 5. Discussing the results

This chapter will interpret and discuss further aspects of the results focusing on details from the dataset, the county's population sizes, and the national average. It will provide a final outlook on the links among, VCA, repression, and the SCS, and finalize by articulating this work's limitations.

5.1. CASM dataset, internet penetration, and urbanization

As pointed on the data segment, Zhang, and Pan (2019) assert that their dataset achieves relatively higher efficiency in capturing CA in rural contexts, at least in comparison to other similar databases. Nonetheless, among other things, my analysis still points to noticeable deficiencies concerning such events. Individual verification showed that, for example, virtually all the counties at the bottom of the list for the VCApc group were rural, whereas this same logic did not apply in the top VCAPpc group. This suggests that the reason behind the higher number of rural counties in the bottom VCApc could be, at least partially, attributed to dataset limitations. Hence the importance of controlling for internet penetration.

5.2. Population size and the national average

Until 2017, the national average of blacklisted individuals per 100.000 inhabitants was only 744.[32] Within the final sample from Table 7, the average number of blacklisted people per 100,000 inhabitants is 97 for the Low VCApc Group, way below the national average. However, even the High VCApc Group of counties had an average number of blacklisted people per 100,000

[32] The number of blacklisted people from 2015 to 03.2018 was circa 9,250,000, and the 2010 census was used to calculate this average.

inhabitants below the national average, in this case, 529.

The reasons for this are also uncertain but it is plausible to consider that, if smaller populations are indeed associated with more blacklists per capita, it would not be surprising that the national average of blacklisted individuals per 100 thousand inhabitants is higher than the ones analyzed here. This could be because all the cities taken for the analysis here have larger populations. In this sense, this may further indicate that there is a negative correlation between population size and the Blacklists. Another possibility might be the fact that the central government includes blacklists from all court levels in this calculation, unfortunately there is no further information available about this.

5.3. Social Credit System = repression?

Since the Court Defaulter's Blacklist is the most advanced part of the SCS, the evidence supporting H1 strengthens the assertion that the SCS is a non-violent repressive tool. It also damages the Chinese Communist Party's official rhetoric about the SCS' focus on enhancing trust and building civic integrity (Wong and Dobson 2019, 220). This does not mean that the system does not aim at addressing those issues too, but repression also seems to be part and parcel of the SCS package.

Yet, it is still unclear how exactly the Court Defaulter's Blacklist would be used to punish protesters. As explained before, since it can be applied to anyone that does not comply with court orders, one could assume that blacklists are being also used as a low-intensity alternative form of coercion before the use of higher intensity forms, like arrest, or killings. Particularly because inflicting the latter would be much more costly in the long run, especially in terms of legitimation (Levitsky and Way 2010, 58; Gerschewski 2013, 21)

A second plausible reason for this discrepancy could be that counties with more VCA are already more prone to have stronger punishment systems in place and, therefore, more people are blacklisted as a consequence. In other words, if people in a county produce more VCAs in a given period, the county administration might be quicker to repress them in the subsequent period. Such logic would further reinforce the assertion that Court Defaulter's Blacklist can be considered a repression tool.

5.4. Limitations

The first and perhaps most severe limitation to this study is the general difficulty to investigate Collective Actions (CAs) in autocratic contexts. People in those environments are already much less prone to rebel in general. According to Zhang and Pan (2019, 48), the sheer threat of suffering violent and non-violent types of repression is a real, and constant fear among the population, and this is usually enough to motivate self-censorship. The authors also point to internet blackouts and website blocks as further difficulties that prevent access to information and could trigger biases that antecede most of the possible measurements one could deploy.

A further difficulty regarding the independent variable Collective Action is the fact this data is collected via Social Media, and this is an inherently more repressed space in an autocratic context. Therefore, the efficiency of the government to delete posts can affect the data collection, particularly in separatistic prone areas (Qin, Strömberg, and Wu 2017). A further related problem is that minority and ethnic conflicts could not be analyzed because the CASM-China dataset has limitations regarding posts in minority languages like in Tibet, and the Uyghur Autonomous Region (Zhang and Pan, 4).

Concerning the dependent variable, the obvious limitation was the reduced sample size. As mentioned in the data segment, the data for the Court Defaulter's Blacklists (and also the other blacklists) lacks in both aggregate and disaggregated forms in all the administrative levels.

Reverse causality is always a risk in the field of contentious politics. Did the VCA cause the blacklists or did the blacklists cause more the VCA? Such concerns are understandable, but they are minimized here particularly because of the timeframe from the VCA and the blacklists. The CASM captures a longer and earlier period (2010 to 2017), while the blacklist system was launched in 2015 and the cases observed here were taken cumulatively from 2015 to 2017. Additionally, since the system is relatively new, its effects might still need time to kick in.

Chapter 6: Conclusion

The claims about the SCS' all-encompassing and controlling nature are scary. Yet, neither the Western media, nor academia have gone far enough to substantiate those claims with empirical evidence. In a closer context, the specialized autocracy literature has investigated the incentives and the forms to which autocrats seek long-term stability using a multitude of integrated tools based on legitimation, co-optation, and repression policies. Referring to the latter, researches point to the deployment of new nonviolent ICT-related ways to coerce the population as a response to dissidence, they range from targeted internet access shutdowns to censorship of undesirable online content.

Nonetheless, given the novelty of the SCS, there is no literature dissecting how it can be conceptualized as one of such tools under an autocratic logic. To fill this gap, I theorized how precisely the SCS could enhance the CCP autocratic stability as a sustainable form of repression, legitimation, and co-optation mechanism all at once. Centering on repression, I examined how the SCS' Court Defaulter Blacklist (CDB) could be considered a new form of nonviolent coercion in response to Collective Actions. Empirically, this thesis aimed at addressing the following research question: do more Violent Collective Actions lead to more Court Defaulter Blacklists?

To tackle it, I used Zhang and Pan (2019)'s Collective Action from Social Media (CASM) dataset to establish Violent Collective Actions per capita (VCApc) as a proxy for the independent variable. I also considered GDP, population, urbanization, and internet penetration as control variables. The data for the dependent variable (the Court Defaulter Blacklists) was difficult to obtain, it was collected manually from annual county courts' performance reports, and focused on the number of CDBs from 2015 to 2017 from each locality. Given those challenges, I implemented a case selection strategy based on the IV, and CVs to build a cross-sectional co-variational study within 6 Chinese counties. Under a small-N logic, this decision was important to focus time and resources on comparing counties that would strengthen the outcomes' inference the most.

The results supported the hypothesis that counties with more VCApcs are associated with more CDBs per capita. Within the selected sample, counties with more VCApc had almost 5 times more CDBs than the ones with fewer VCApc, even with all the selected counties being poorer, rural, largely populated, and with higher levels of internet penetration. Additional tests to enhance external validity also showed similar results in a random sample, and among richer counties. Nonetheless, a probe with both richer and more urban counties showed contrary evidence to this work's hypothesis, possibly because the only county marching all the case selection's criteria for the low VCApc group was the Rongcheng county, widely recognized as an SCS model county, and might therefore be an outlier.

Further research could explore other consequences of the blacklists over a longer period: could they undermine future Collective Action movements in China? Additionally, the specialized autocracy literature could investigate empirically how other SCS' components might enhance legitimation or co-optation mechanisms In this sense, since the public entities administrators can be personally blacklisted for their agencies' poor management, this could affect governance indicators. What's more, the local municipal point systems are very diverse and might have great influence over people's behavior, thus there could be a myriad of consequences in the population's attitudes that could be studied.

This fits into a wider debate about the dynamics of big data and surveillance as new forms of autocratic repression to autocratic stabilization. Furthermore, this study can be considered a plausibility probe to pave the way for researchers to test if the relation between Collective Actions and the SCS holds in a large-N investigation. Ultimately, this work also helps shed light on the SCS' broader implication. Nationally, for example, it can serve as a baseline to contextualize the recent leaks pointing to the use of a similar credit system in Chinese prisons (Buckley 2019), and the instrumentalization of big data-driven algorithms to identify and arrest individuals belonging to Uyghur Muslim minorities (Allen-Ebrahimian 2019). Internationally, it can help red flag the creation of similar initiatives like the ones planned for India (Shahin and Zheng 2020). Finally, while one needs to be careful about assuming what might have been the ruler's true intent for a policy, it is perfectly practical to examine the factors that might have shaped its causes and effects.

Note: The links to the original Court Defaulter Blacklists county court reports, the case selection tables, and further data references can be found in the following link: https://docs.google.com/spreadsheets/d/1Yjp9dp-VXPJowfsMEbyeoilmdSEXN0gxtVSeH1ZAEMA/edit?usp=sharing

Chapter 7: Bibliography

1. Allen-Ebrahimian, Bethany. 2019. "Exposed: China's Operating Manuals for Mass Internment and Arrest by Algorithm." *International Consortium of Investigative Journalists*.
2. Almeida, Paul D. 2003. "Opportunity Organizations and Threat-Induced Contention: Protest Waves in Authoritarian Settings." *American Journal of Sociology* 109 (2): 345–400. https://doi.org/10.1086/378395.
3. Barton, Allen H. 1955. "The Concept of Property-Space in Social Research." *The Language of Social Research*, 40–53.
4. Berman, Lex. 2017. "China Population by County (1999)." Harvard Dataverse. https://doi.org/10.7910/DVN/EOH3FV.
5. Bernstein, Thomas P. 2004. "Unrest in Rural China: A 2003 Assessment," August. https://escholarship.org/uc/item/1318d3rx.
6. Blatter, Joachim, and Markus Haverland. 2012. *Designing Case Studies*. London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137016669.
7. Brumberg, Daniel. 2002. "Democratization in the Arab World? The Trap of Liberalized Autocracy." *Journal of Democracy* 13 (4): 56–68.
8. Buckley, Chris. 2019. "China's Prisons Swell After Deluge of Arrests Engulfs Muslims." *The New York Times*, August 31, 2019, sec. World. https://www.nytimes.com/2019/08/31/world/asia/xinjiang-china-uighurs-prisons.html.
9. Cai, Yongshun. 2010. *Collective Resistance in China: Why Popular Protests Succeed Or Fail*. Stanford University Press.
10. Chen, Yongxi, and Anne S. Y. Cheung. 2017. "The Transparent Self Under Big Data Profiling: Privacy and Chinese Legislation on the Social Credit System." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2992537.
11. Chen, J., Pan, J. and Xu, Y., 2016. Sources of authoritarian responsiveness: A field experiment in China. American Journal of Political Science, 60(2), pp.383-400.
12. China's State Council. 2014. "State Council Notice Concerning the Issuance of the Planning Outline for the Construction of a Social Credit System (2014-2020)." June 14, 2014. https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/.

13. Chinese Academy Of Surveying And Mapping. 1997. "China Dimensions Data Collection: GuoBiao (GB) Codes for the Administrative Divisions of the Peoples Republic of China." Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4KK98PS.

14. Dai, Xin. 2018. "Toward a Reputation State: The Social Credit System Project of China." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3193577.

15. Dasgupta, Susmita, Somik Lall, and David Wheeler. 2005. "Policy Reform, Economic Growth and the Digital Divide." *Oxford Development Studies* 33 (2): 229–43. https://doi.org/10.1080/13600810500137889.

16. Davenport, Christian. 2007a. "State Repression and Political Order." *Annual Review of Political Science* 10 (1): 1–23. https://doi.org/10.1146/annurev.polisci.10.101405.143216.

17. ———. 2007b. "State Repression and the Tyrannical Peace." *Journal of Peace Research* 44 (4): 485–504. https://doi.org/10.1177/0022343307078940.

18. Dimitrov, Martin, and Zhu Zhang. 2017. "Patterns of Protest Activity in China."

19. Distelhorst, Greg, and Yue Hou. 2017. "Constituency service under nondemocratic rule: Evidence from China." The Journal of Politics 79, no. 3: 1024-1040.

20. Dukalskis, Alexander, and Johannes Gerschewski. 2018. "Adapting or Freezing? Ideological Reactions of Communist Regimes to a Post-Communist World." *Government and Opposition*, November, 1–22. https://doi.org/10.1017/gov.2018.40.

21. Eckstein, Harry. 1975. "Case Study and Theory in Political Science." In *Handbook of Political Science, Vol. 7*, edited by Fred I. Greenstein and Nelson W. Polsby, 79–138. Reading, MA: Addison-Wesley Press. http://publishing.cdlib.org/ucpressebooks/view?docId=ft0k40037v&chunk.id=d0e2292&toc.depth=1&toc.id=d0e2292&brand=ucpress.

22. Engelmann, Severin, Mo Chen, Felix Fischer, Ching-yu Kao, and Jens Grossklags. 2019. "Clear Sanctions, Vague Rewards: How China's Social Credit System Currently Defines 'Good' and 'Bad' Behavior." In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, 69–78. Atlanta, GA, USA: ACM Press. https://doi.org/10.1145/3287560.3287585.

23. Foreign Policy. 2018. "China Is Implementing a Massive Plan to Rank Its Citizens, and Many of Them Want In." *Foreign Policy* (blog). 2018. https://foreignpolicy.com/2018/04/03/life-inside-chinas-social-credit-laboratory/.

24. Gan, Nectar. 2019. "The Complex Reality of China's Soical Credit System: Hi-Tech Dystopian Plot or Low-Key Incentive Scheme?"." South China Morning Post. 2019. https://www.scmp.com/news/china/politics/article/2185303/hi-tech-dystopia-or-low-key-incentive-scheme-complex-reality.

25. Gerschewski, Johannes. 2013. "The Three Pillars of Stability: Legitimation, Repression, and Co-Optation in Autocratic Regimes." *Democratization* 20 (1): 13–38. https://doi.org/10.1080/13510347.2013.738860.

26. Goebel, Christian. 2017. *Social Unrest in China: A Bird's Eye Perspective*.

27. Gohdes, Anita R. 2015. "Pulling the Plug: Network Disruptions and Violence in Civil Conflict." *Journal of Peace Research* 52 (3): 352–67. https://doi.org/10.1177/0022343314551398.

28. Goldstein, Robert Justin. 1978. *Political Repression in Modern America from 1870 to the Present*. GK Hall & Company.

29. Goldstone, Jack Andrew, and Charles Tilly. 2001. "Threat (and Opportunity): Popular Action and State Response in the Dynamics of Contentious Action." In *Silence and Voice in the Study of Contentious Politics*, 179.

30. Gurr, Ted Robert. 1986. "Persisting Patterns of Repression and Rebellion: Foundations for a General Theory of Political Coercion." *Persistent Patterns and Emergent Structures in a Waning Century*, 149–68.

31. Hassanpour, Navid. 2014. "Media Disruption and Revolutionary Unrest: Evidence From Mubarak's Quasi-Experiment." *Political Communication* 31 (1): 1–24. https://doi.org/10.1080/10584609.2012.737439.

32. Hess, Steve. 2010. "Nail-Houses, Land Rights, and Frames of Injustice on China's Protest Landscape." *Asian Survey* 50 (5): 908–26. https://doi.org/10.1525/as.2010.50.5.908.

33. Hoffman, S. 2017. "Managing the State: Social Credit, Surveillance and the CCP's Plan for China." *China Brief* 17 (11): 21–27.

34. Horsley, Jamie. 2018. "China's Orwellian Social Credit Score Isn't Real." *Foreign Policy* (blog). 2018. https://foreignpolicy.com/2018/11/16/chinas-orwellian-social-credit-score-isnt-real/.

35. Hurst, William. 2004. "Understanding Contentious Collective Action by Chinese Laid-off Workers: The Importance of Regional Political Economy." *Studies in Comparative*

International Development 39 (2): 94–120. https://doi.org/10.1007/BF02686279.

36. Institute of Social Science Survey, Peking University. 2017. "China Family Panel Studies (2010) Baseline Survey." Harvard Dataverse. https://doi.org/10.7910/DVN/PF7FSS.

37. Johnston, Hank. 2012. "State Violence and Oppositional Protest in High-Capacity Authoritarian Regimes" 6: 21.

38. 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 (02): 326–43. https://doi.org/10.1017/S0003055413000014.

39. Kostka, Genia, and Lukas Antoine. 2018. "Fostering Model Citizenship: Behavioral Responses to China's Emerging Social Credit Systems." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3305724.

40. Levitsky, Steven, and Lucan Way. 2002. "The Rise of Competitive Authoritarianism." *Journal of Democracy* 13 (2): 51–65. https://doi.org/10.1353/jod.2002.0026.

41. Lijphart, Arend. 1975. "II. The Comparable-Cases Strategy in Comparative Research." *Comparative Political Studies* 8 (2): 158–77. https://doi.org/10.1177/001041407500800203.

42. Lin, Fen, Tsan-Kuo Chang, and Xinzhi Zhang. 2015. "After the Spillover Effect: News Flows and Power Relations in Chinese Mainstream Media." *Asian Journal of Communication* 25 (3): 235–54. https://doi.org/10.1080/01292986.2014.955859.

43. Liu, Chuncheng. 2019. "Multiple Social Credit Systems in China," July. https://doi.org/10.31235/osf.io/v9frs.

44. Mac Síthigh, Daithí, and Mathias Siems. 2019. "The Chinese Social Credit System : A Model for Other Countries?" Working Paper. http://cadmus.eui.eu//handle/1814/60424.

45. Mckenzie, Lachlan, and Katharina L. Meissner. 2016. "Human Rights Conditionality in European Union Trade Negotiations: The Case of the EU-Singapore FTA: Conditionality in EU Trade Negotiations." *JCMS: Journal of Common Market Studies* 55 (4): 832–49. https://doi.org/10.1111/jcms.12522.

46. Meng, Tianguang, Jennifer Pan, and Ping Yang. 2017. "Conditional Receptivity to Citizen Participation: Evidence From a Survey Experiment in China." *Comparative Political Studies* 50 (4): 399–433. https://doi.org/10.1177/0010414014556212.

47. Ngai, Pun, and Lu Huilin. 2010. "Unfinished Proletarianization: Self, Anger, and Class

Action among the Second Generation of Peasant-Workers in Present-Day China." *Modern China* 36 (5): 493–519. https://doi.org/10.1177/0097700410373576.

48. O'Brien, Kevin J., and Yanhua Deng. 2015. "Repression Backfires: Tactical Radicalization and Protest Spectacle in Rural China." *Journal of Contemporary China* 24 (93): 457–70. https://doi.org/10.1080/10670564.2014.953849.
49. Ohlberg, Mareike, Shazeda Ahmed, and Bertram Lang. 2017. "CENTRAL PLANNING, LOCAL EXPERIMENTS," 15.
50. Pence, Mike. 2018. "Remarks by Vice President Pence on the Administration's Policy Toward China." The White House. 2018. https://www.whitehouse.gov/briefings-statements/remarks-vice-president-pence-administrations-policy-toward-china/.
51. Planet Money NPR. 2018a. "#871: Blacklisted In China : NPR." 2018. https://www.npr.org/templates/transcript/transcript.php?storyId=661151404.
52. ———. 2018b. "China's Social Credit System : NPR." 2018. https://www.npr.org/templates/transcript/transcript.php?storyId=655914994.
53. Pu, Qiongyou, and Stephen J. Scanlan. 2012. "Communicating Injustice?" *Information, Communication & Society* 15 (4): 572–90. https://doi.org/10.1080/1369118X.2012.665937.
54. Qin, Bei, David Strömberg, and Yanhui Wu. 2017. "Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda," 32.
55. Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics*." *American Political Science Review* 64 (4): 1033–53. https://doi.org/10.2307/1958356.
56. Schaefer, Kendra, and Ether Yin. 2019. "What Is Social Credit? - Trivium Social Credit - Tracking China's Social Credit System." *Trivium Social Credit* (blog). 2019. http://socialcredit.triviumchina.com/what-is-social-credit/.
57. Selden, Mark, and Elizabeth J Perry. 2010. "Introduction: Reform, Conflict and Resistance in Contemporary China." In *Chinese Society*, 19–48. Routledge.
58. Shahin, Saif, and Pei Zheng. 2020. "Big Data and the Illusion of Choice: Comparing the Evolution of India's Aadhaar and China's Social Credit System as Technosocial Discourses." *Social Science Computer Review* 38 (1): 25–41.
59. Shorey, S., and P. Howard. 2016. "Automation, Big Data and Politics: A Research Review." *International Journal of Communication* 10. https://ora.ox.ac.uk/objects/uuid:252c907d-9357-47bc-b3be-7f2e4ba1fddd.

60. Slater, Dan, and Sofia Fenner. 2011. "STATE POWER AND STAYING POWER: INFRASTRUCTURAL MECHANISMS AND AUTHORITARIAN DURABILITY." *Journal of International Affairs* 65 (1): 15–29.

61. Soros, George. 2019. *In Defense of Open Society*. Hachette UK.

62. Supreme People's Court of China. 2017. "More than 170,000 Defaulters Banned from Senior Management Positions in China." 2017. http://english.court.gov.cn/2017-11/21/content_34809918.htm.

63. Tarrow, Sidney, and Charles Tilly. 2007. "Contentious Politics and Social Movements." In *The Oxford Handbook of Comparative Politics*.

64. Wong, Karen Li Xan, and Amy Shields Dobson. 2019. "We're Just Data: Exploring China's Social Credit System in Relation to Digital Platform Ratings Cultures in Westernised Democracies." *Global Media and China* 4 (2): 220–32. https://doi.org/10.1177/2059436419856090.

65. Wunnava, Phanindra V., and Daniel B. Leiter. 2009. "Determinants of Intercountry Internet Diffusion Rates." *American Journal of Economics and Sociology* 68 (2): 413–26. https://doi.org/10.1111/j.1536-7150.2009.00634.x.

66. Zhang, Han, and Jennifer Pan. 2019. "CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media:" *Sociological Methodology*, July. https://doi.org/10.1177/0081175019860244.

67. Zhang, Hongri. 2017. "More than 100 Local Governments Were Classified into Discredited Judgement Debtor List." *Guancha. Cn*.