From Internet to Social Safety Net: The Policy Consequences of Online Participation in China

Junyan Jiang†  Tianguang Meng‡  Qing Zhang§
University of Pennsylvania  Tsinghua University  Columbia University

Abstract

The internet is widely touted for its potential to reduce inequality in political influence among citizens, but research focusing on advanced democracies has yielded mixed results. We provide new evidence on the internet’s equalizing impact by studying how political participation via a major national online petition forum affects local government policies in China, an authoritarian regime with the world’s largest internet-using population. Content analysis of over 900,000 petitions reveals that the forum was predominantly used by citizens from rural areas and low-income groups. Linking geographic and temporal variations in petition volumes to a new dataset of government policy priorities, we further show that increases in online participation led city governments to place greater emphasis on social welfare issues, which tend to disproportionately benefit the poor. Additional analyses suggest that online participation induces policy changes by both raising the authority’s awareness of disadvantaged citizens and amplifying the threats of collective actions.

*Earlier versions of the paper have been presented at the Annual Meetings of the American Political Science Association. For helpful discussion and comments, we thank Douglas Almond, David Blei, Anthony Fowler, Amit Khandelwal, Wojciech Kopczuk, Pierre Landry, Feng Li, Christian Goebel, Dan Hopkins, Ying Lu, W. Bentley MacLeod, Suresh Naidu, Jennifer Pan, Molley Roberts, Weiyi Shi, Victor Shih, Susan Shirk, Elizabeth Stoycheff, Rory Truex, Miguel Urquiola, Eric Verhoogen, Xiaoyu Wang, Alfred Wu, Yiqing Xu, Dali Yang, and seminar participants at University of Amsterdam, Columbia, Hong Kong Institute of Education, and UC San Diego. We are grateful to Manfred Elfstrom for generously sharing his labor protest data.

† Postdoctoral Fellow, Center for the Study of Contemporary China, junyanji@sas.upenn.edu.
‡ Associate Professor, Department of Political Science, maxmeng@mail.tsinghua.edu.cn.
§ PhD Candidate, Department of Economics, qz2208@columbia.edu.
1 Introduction

Not all citizens enjoy an equal voice in politics. Even in a democratic system, where political equality is a guiding principle, citizens with different social and class backgrounds often participate in politics at radically different rates (Lijphart 1997). A extensive body of research has shown that individuals with higher socioeconomic status are more likely to turnout to vote, make campaign contributions, and contact governments via various means (Kasara 2007; Schlozman, Verba and Brady 2012; Verba and Nie 1972), and their activism often translates into greater de facto influence over the political and policy-making process (Gilens 2012; Griffin and Newman 2005). If anything, this influence gap is likely to be even larger in authoritarian regimes, which by definition concentrate power within a smaller group of elites (Bueno de Mesquita et al. 2003; Linz 2000). Inequality in policy influence is believed to have not only troubling normative implications (at least for democracies) but also important practical consequences. Recent research has shown that interventions aimed at lowering the cost of electoral participation can produce more egalitarian distributive and policy outcomes that favor lower-class citizens (Fowler 2013; Fujiwara 2015).

As one of the most influential technological inventions of the 20th century, the internet has been repeatedly touted for its potential to bring greater equality into politics. Many scholars and policy practitioners believe that the internet offers a direct and inexpensive way for marginalized groups to access political information, contact public officials, organize collective actions (Norris 1999), and thereby give “new voice to people who’ve felt voiceless” (Gillmor 2004). However, existing research on the political consequences of the internet, which focuses primarily on advanced democracies (the United States in particular), has not yet found convincing evidence for such an equalizing effect. A number of survey-based studies have shown that, instead of helping to level the playing field in politics, the internet tends to reinforce and even exacerbate the existing participatory gap across classes and socioeconomic status (Jennings and Zeitner 2003; Norris 2001; Schlozman, Verba and Brady 2012, 483-533) and amplify elite voices (Hindman 2009; Margolis and Resnick 2000). More importantly, while some recent studies do find evidence that internet usage may increase civic engagement for certain underprivileged groups (Cohen and Luttig 2016;
Smith 2014), there is not yet any systematic evidence on whether and how changes in participation patterns can alter substantive policy outcomes.

In this article, we provide new evidence on the political consequences of the internet by studying how internet-based participation shapes government policies in China. As an authoritarian regime that hosts the world’s largest internet-using population,¹ the Chinese case is not only practically important but also theoretically interesting, as existing theories about internet and authoritarian politics gives us contrasting expectations. On the one hand, conventional theories of authoritarian politics would suggest that the effect of online participation on government policies is likely to be minimal, as there is no electoral pressure for government to be responsive to popular demands. On the other hand, limited avenues for traditional participation, coupled with a large number of internet users, can also mean that online participation will have an unusually large effect on government policy, especially when compared to advanced democracies, where citizens already enjoy a full menu of offline methods to make their voices heard.

More specifically, this article studies the patterns and policy ramifications of mass participation through the Local Leader Message Board (LLMB), a large, national-level online petition forum that allows citizens to directly register complaints to leading party and government officials in their localities. Compared to the more traditional methods of filing petitions by mail or in person, this online forum offers an unusually cheap and transparent alternative for citizens to communicate their concerns to the political authority. We expect this method to be especially attractive to lower-class citizens, who tend to be underrepresented in the more traditional political institutions and are usually most sensitive to changes in participation cost. Based on a large body of research suggesting that low-income groups tend to hold stronger preference for redistribution than high-income individuals (Acemoglu and Robinson 2005; Alesina and Ferrara 2005; Meltzer and Richard 1981), we further hypothesize that expanded online participation is likely to shift government policies towards placing greater emphasis on social welfare and redistributive issues, as voices of lower-class citizens become better heard.

¹According to China Internet Network Information Center (CNNIC), the number of internet users in China is estimated to be over 710 million as of July 2016.
Our empirical analysis begins with an examination of the socioeconomic backgrounds of forum participants. We manually code posters’ occupations for over 70,000 petition, and then use a new non-parametric content analysis method developed by Hopkins and King (2010) to estimate the distribution of occupations within the entire body of over 900,000 petitions based on linguistic similarity. We find that, compared to nationally representative samples, citizens with lower class backgrounds, such as rural residents and low-income urban workers, have a disproportionately large presence in the petition forum while better-off groups, such as party cadres, government employees and private entrepreneurs, tend to be underrepresented.

We then investigate how expanded online participation affected the substantive policy priorities of local governments. To construct a consistent measure of government policies that can be compared across time and space, we apply machine-learning techniques on over 4,400 Government Work Reports (GWRs), which are comprehensive policy blueprints published by the Chinese government, and estimate the relative proportions for different “topics” (i.e., clusters of words) as a measure of government policy priorities. Using a series of fixed-effects models, we confirm the hypothesis that increases in participation through the LLMB have a measurable impact on government policies. More specifically, we find that cities that received a larger number of online petitions from local citizens tended to devote a significantly higher proportion of the government reports in the following year to a topic on social welfare, which broadly includes issues such as unemployment benefits, low-income/disability subsidies, medical care, and pension coverage. We also find evidence that these changes in stated policy priorities were not merely cheap talk, but were instead accompanied by meaningful changes in other outcomes, including more legislative activities and fiscal spending on welfare-related issues.

We also evaluate three potential mechanisms by which online petitions affect policy changes: (1) information revelation, (2) amplifying the threats of collective actions, and (3) facilitating top-down monitoring by political superiors. Our analysis suggests that petitions that are filed at local decision makers’ own message boards and those during periods with high government attention is especially likely to induce policy changes; these patterns provide support to the information
revelation mechanism. We also find that, consistent with the collective action threat mechanism, increases in offline protest can induce more frequent replies to individual requests and that the marginal impact of online petitioning on policy is larger in localities with greater collective action potential. By contrast, we show that senior leaders’ interest in the petition forum does not significantly alter policy responsiveness of lower-level governments, suggesting that better top-down monitoring may not be a main mechanism that drives policy changes.

By offering, to our knowledge, the first systematic assessment of the policy consequences of internet-based participation, this study is related first and foremost to a large literature on how the internet, or advancement in information communication technologies (ICTs) in general, shapes political activities and quality of governance. Existing research has found that new ICTs can help strengthen government accountability (Besley and Burgess 2002; Rezaee, Hasanain and Khan 2015) and reduce corruption (Bailard 2009), but may also facilitate violent collective actions (Pierskalla and Hollenbach 2013) and increase ideological polarization (Lelkes, Sood and Iyengar 2015). Most closely related to our work here, recent studies by Grossman, Humphreys and Sacramone-Lutz (2014, 2016) use field experiments to investigate the effect of Short Message Services (SMS) on interest articulation of marginalized populations in Uganda. While the authors find a significant flattening effect of SMS in an earlier pilot (2014), they find no such effect from a follow-up, nationwide experiment (2016). Few existing studies, however, have gone a step further to investigate whether and how the new, ICT-based participation can produce meaningful behavioral changes on the part of the government. Our study not only confirms ICT’s potential to mitigate participatory inequality among citizens, but also provides further evidence that technology-induced closing of participation gaps can indeed have substantive effects on local governments’ policy priorities even in a system that lacks meaningful electoral or representative institutions.

Moreover, our study also contributes to a nascent but rapidly growing literature on responsiveness in authoritarian regimes. Several prior studies have shown that in some autocracies, and China in particular, inquiries and demands made by individual citizens receive replies from gov-
ernments at a comparable rate as in democracies (Chen, Pan and Xu 2016; Distelhorst and Hou Forthcoming), and that individual politicians are generally receptive to citizen opinions (Manion 2016; Meng, Pan and Yang 2014; Truex 2016). The existing literature, however, has not yet provided any systematic evidence as to whether the preferences and demands from citizens are actually being taken into account in government policy making—a much more fundamental criterion of political responsiveness according to classical writings on this concept (Dahl 1971; Miller and Stokes 1963). By showing that online participation by Chinese citizens can indeed collectively impact local government policies, our study provides more definitive evidence on the existence of a form of authoritarian responsiveness that conforms with its conventional usage in the democratic context.

2 Participation Bias in Authoritarian Regimes and the Equalizing Role of Internet

Traditional theories of authoritarian regimes posit that there are few meaningful ways for citizens to participate in politics under those regimes (Linz 2000). This view, however, has been increasingly challenged. Recent studies have argued that it may be in the survival interest of autocratic rulers to allow limited participation by citizens as a way to collect information and preempt more violent uprisings. Commonly studied participation mechanisms in authoritarian regimes include semi-competitive elections (Blaydes 2010; Levitsky and Way 2010), petitions (Dimitrov 2014) and/or a relatively independent media to which citizens can whistleblow wrongdoings of lower-level officials (Egorov, Guriev and Sonin 2009).

Little attention, however, has been paid to inequality in citizens’ ability to access and utilize participatory institutions in nondemocratic systems. Existing research on democracies suggest that citizens with different levels of interests, skills, and resources may participate in politics at different rates (e.g., Brady, Verba and Schlozman 1995), and some of these factors may also play a role in shaping citizens’ propensity to participate in nondemocracies. Aside from the variations from the
demand side, autocrats may also deliberately supply their attention in a way that favors certain
groups over others. One classical manifestation of such a tendency is the so-called urban bias
(Lipton 1977; Wallace 2013), whereby citizens living in major cities receive disproportionately
more attention from autocrats than do rural residents because of the former’s greater economic and
political significance. More generally, the selectorate theory developed by Bueno de Mesquita et al.
(2003) also argues that authoritarian leaders’ survival depends first and foremost on maintaining
support from a winning coalition among members of the selectorate, implying that they will pay
greater attention to the interest and demands of their elite supporters in making policies. Follow-
up empirical research has shown that regimes with small coalitions tend to under-provide public
goods and implement more regressive policies that benefit the rich and the powerful (e.g., Bueno de
Mesquita, Downs and Smith 2017; Morrow et al. 2008).

Democratization is often seen as a means to rectify the political inequality in autocratic systems
(Acemoglu and Robinson 2005; Meltzer and Richard 1981), although the empirical evidence on
the equalizing impact of democratization still remains mixed (Gallagher and Hanson 2009). We
argue that advancement in ICTs—and the internet in particular—offers another way to reduce the
de facto influence gap among citizens under autocratic systems without inducing systemic regime
changes. By lowering the physical barriers and the financial costs associated with communication,
the internet helps improve citizens’ ability to communicate their grievances to authorities and to or-
ganize collective actions among themselves (Bennett and Segerberg 2015). The reduction in com-
munication and organizing costs can be especially appealing to the poor and the underprivileged,
as they often lack the necessary resources or capacity to effectively utilize the more traditional par-
ticipatory institutions (Gibson, Lusoli and Ward 2005; Krueger 2002). Moreover, recent research
also suggests that governments’ inclusion measures often generate more positive psychological
reactions from less-educated and more marginalized groups, because these groups tend to have
relatively low standards/expectations for how the government should behave (Truex 2017). When
being presented with a new and inexpensive channel of participation, therefore, lower-class citi-
zens are more likely to experience a “satisfaction boost” with the government, which may further
increase their willingness to engage with the authority via the new channel.

It is also important to note that compared to democratic systems, authoritarian regimes may be especially attentive to demands made via these new channels due to the underdevelopment of the more traditional forms of participation. The absence of strong representative institutions and regular, competitive elections implies that authoritarian governments often have difficulty acquiring accurate and complete information about citizens’ demands and opinions through the traditional, institutionalized channels (Kuran 1997; Lorentzen 2013); this can motivate them to place greater weight on information conveyed through alternative mechanisms, including the internet, when making political and policy decisions. Moreover, online discussions are sometimes harbingers of more disruptive offline collective actions, which typically pose a much greater existential threat to autocratic than democratic governments. While sometimes threats can be neutralized through repressive measures that place restrictions on the scope and content of internet discussions (King, Pan and Roberts 2013), in certain cases repression can be too costly to carry out or simply ineffective, and autocrats may be motivated to pursue a more pacifist strategy that involves making substantive policy concessions in response to citizens’ demands.

3 Traditional and Internet-based Participation in China

Although Chinese citizens do not enjoy the right to elect their government officials, prior research has shown that the regime does have several functioning non-electoral channels through which citizens can make their voices heard (Shi 1997; Tang 2016). Officially sanctioned channels include participating in elections for community/grassroots leaders, filing petitions to party and government agencies, attending government-organized hearings, and contacting the media, local officials, or People’s Congress deputies (Manion 2016; O’Brien and Li 2006; Shi 1997; Truex 2016). Some recent research also suggests that certain types of protests and collective actions may be implicitly tolerated by the higher-level authority as a way to gather information about its subordinates (e.g., Lorentzen 2013). While there is growing evidence that the presence of these participation channels
has made the regime more inclusive and contributed to its ability to deliver effective governance (Stromseth, Malesky and Gueorguiev 2017), an important but often overlooked caveat is these channels are often not utilized equally by citizens from all social groups. Empirical research on political participation in China has consistently documented that individuals who are wealthier, more educated, and closer to the regime (i.e., with affiliation with the party, the government or public organizations) are more likely to participate in competitive local elections, communicate their issues to the government, and/or reflect their personal grievances to the media (Guo 2007; Landry, Davis and Wang 2010; Tsai and Xu 2013). The differential rate of utilization is a result of not only the difference in information access and resource endowment across classes, but also some highly discriminatory designs of the traditional institutions. Participation in grassroots elections, for instance, is often limited to those officially registered as local residents (i.e., with a local *hukou*) and typically excludes migrant workers, who constitute a sizable share of the local population in many affluent regions. Similarly, quotas for deputies in national and legislative bodies are often allocated in favor of the wealthy class, such as private entrepreneurs, intellectuals, professionals, and urban residents in general (Truex 2016).\(^2\)

The rapid expansion of the internet in China since the 1990s and especially during the first decade of the 21st century has brought profound changes to the way citizens and governments interact. The emergence of online forums, weblogs, and many other types of virtual communities have facilitated lively discussion of public affairs and provided citizens with new channels to articulate their personal interests and problems in a way that draws broad attention (Yang 2009). At the same time, seeing the internet’s potential as an important outlet for publicity and a means to improve administrative effectiveness, the political authority has also undertaken systematic efforts to expand its online presence. The word “E-government (*dianzi zhengfu* 电子政府)” first appeared in a major party policy document at the 16th Party Congress in 2002, and the State Council subsequently undertook a series of measures to promote the development of government websites, with

\(^2\)According to the law, urban residents are entitled to four times as many deputies as rural residents. See *Election Law on National and Local People’s Congress* [http://www.people.com.cn/GB/14576/28320/39838/39842/2947944.html](http://www.people.com.cn/GB/14576/28320/39838/39842/2947944.html).
a special emphasis on information availability and ease of public access. As of 2008, almost all local governments at or above the city-level had created their own website, and the majority of the website contained at least some function for citizens to communicate their views and register grievances, albeit with varying degrees of effectiveness.

The question of how the emergence of the online medium has altered the composition of citizens participating in politics, however, has not yet been adequately answered in the extant research. In the rest of this section, we provide some preliminary evidence on this issue by drawing on individual-level data from the China Citizenship Attitude Survey (CCAS), a nationally representative survey conducted by Peking University in 2008 and 2009. The survey contains a battery of questions on respondents’ past political participation. We construct two binary variables: *Traditional Participation*, which indicates whether a respondent has ever used any of the conventional, offline methods to contact the government, and *Internet Participation*, which asks whether the respondent has ever “participated or expressed opinions in an online platform on issues related to politics.” We then explore how (logged) family income, a proxy for socioeconomic status, affects respondents’ propensity to participate. Figure 1 plots three simple bivariate relationships using locally weighted regression. There appears to be a strong, positive relationship between log family income and participation through the traditional channels (the blue solid line). This relationship, however, is much weaker for the internet-based channel (the yellow dotted line), and effectively disappears when we further restrict to the sample of respondents who use the internet on a weekly basis (the red dashed line).

These contrasting patterns suggest that there exists quite substantial de facto inequality in the traditional participation channel, but such inequality is much less severe in the internet-based ones,

---

3 More specifically, this variable takes the value of 1 if the respondent had used at least one of the following channels in the past: (1) attending government meeting, (2) making appeals to the upper-level government, and (3) contacting media, local cadres, or People’s Congress deputies to express their views (questions f10a, f10b, and f10c).

4 This is item F10e in the questionnaire.

5 Estimating this relationship using linear regression yields a coefficient of $\hat{\beta} = 0.04$ ($p$ value $< 0.001$), controlling for age, education, and year and city fixed-effects.

6 $\hat{\beta} = 0.0017$ ($p$ value = 0.621)

7 More specifically, this sample includes respondents who report to have used the internet at least once in the last week, based on question c1_e. The coefficient estimates $\hat{\beta} = 0.0264$ ($p$ value = 0.236)
in which the rich and the poor appear to be participating at comparable rates. As political participation in the cyberspace expands, therefore, we shall expect more voices of lower-class citizens to be brought into the political process. When the government is truly responsive to citizen demands, then expanded online participation are likely to make its policies more aligned with the interests of the poor and disadvantaged. More specifically, based on a large body of political economy research (e.g., Acemoglu and Robinson 2005; Meltzer and Richard 1981; Hibbs, Rivers and Vasilatos 1982), we hypothesize that the direction of policy shift will be toward greater emphasis on redistribution and social welfare, as lower-class citizens tend to benefit disproportionately from public goods and government-sponsored welfare programs.8

4 The Local Leader Message Board

The Local Leader Message Board (LLMB http://liuyan.people.com.cn/) was created and operated by People’s Daily Online (人民网), the official website of the central state media. The forum was formally launched in 2008, after which it quickly became a popular venue for ordinary citizens to file complaints directly to local officials. Inspired by this model, a few localities later also set up their own platforms to receive petitions, but the LLMB remains up to this day the largest and the only national petition forum in China. As of August 2016, the forum has received over 900,000 petitions filed by citizens from all Chinese provinces (except Taiwan). Among them, over 550,000, or 60 percent, have received replies from local authorities.

Figure 2 provides a snapshot of the LLMB’s web interface.9 For every regional administration at the county level or above (i.e., 33 provinces, 333 cities, and over 2,000 counties), the website contains two message boards for the head of the party and the head of the government, respectively. Within each message board, all messages are listed in a chronological order starting with the most recent ones; government replies, if any, will be attached to the bottom of the original messages.

---

8 We provide additional survey evidence on the diverging preferences between the rich and the poor in the Online Appendix.

9 A smartphone application of the LLMB also recently became available (https://itunes.apple.com/cn/app/id828180925?mt=8). The interface is slightly different but retains all the essential functions.
Participants can leave messages either anonymously or with a registered account. Registration requires only a functional cell phone number or email address, and will allow the users to customize the names that appear on their messages. After a petition is filed and before it is publicly posted, the website will undertake a quick check to ensure that the message contains no vulgar language or any politically sensitive content. According to LLMB staffs that we interviewed, censorship is only limited to posts that contain explicit comments or criticisms about top central leaders.\textsuperscript{10}

Insofar as policy responsiveness is concerned, several features of the LLMB deserve special mentioning. First, in contrast to more locally operated websites, in which the local authorities have a direct control over the content to be displayed, the LLMB is operated by a central agency, which typically has little direct interest in helping local authorities cover up their own problems. This means that local governments will not be able to simply bury any undesirable issues through censorship. Additionally, the website also employs several measures to increase the publicity of petitions and put pressure on local government to respond. The names of incumbent local leaders, for example, are printed in boldface at the top of their own message boards, along with a few summary statistics, such as the cumulative number of petitions received and replied. All petitions and government replies, once posted, are publicly visible to all users; this not only makes it easier for local residents with common grievances to find each other and organize but can also help local leaders’ political superiors to directly learn about citizens’ opinions about their subordinates. Occasionally, the LLMB’s own news team would also actively look for newsworthy materials from the petitions and conduct follow-up investigations. All these features—the independence of the operator, the transparent design of the website, and the high publicity of the platform—may give local officials an incentive to be attentive to citizen demands made on the forum.

Anecdotal evidence suggests that, at least on surface, local governments do appear to take petitions at the LLMB seriously: Since June 2008, 19 provinces have set up specialized agencies and standard work routines to process LLMB petitions and to regularly communicate important information from these petitions to the local leadership.\textsuperscript{11} Over 50 provincial party secretaries

\textsuperscript{10}Personal interviews in Beijing, December 2016.
\textsuperscript{11}Typically, a special office will be created to handle all the online petitions and allocate them to different depart-
and governors, the highest-ranking regional leaders in China, have personally responded to citizen petitions via the LLMB, along with many more city- and county-level leaders.

5 Who Participates?

We scraped the content of all the publicly available petitions from the website, along with a large set of available auxiliary information for each petition, including its subject matter, time of posting, the identity of the leader to whom the petition was directed, whether a government reply has been made, and so on. Several rounds of scraping were conducted between 2014 and 2016, and the outputs were combined into a single dataset with duplicated records removed.¹²

To understand the implications of LLMB for the landscape of political participation in China, we begin by providing a descriptive analysis of the distribution of occupational backgrounds among LLMB users. We do so by first manually coding posters’ occupations for a subset of petitions and then use it to infer the overall occupation distribution in the entire body of petitions using the semi-automated content analysis method developed by Hopkins and King (2010).¹³ Our key assumption here is that individuals from the same occupational backgrounds may share similar linguistic patterns. We believe that this assumption is likely to hold because a lot of the common grievances raised on the forum, such as land expropriation, contract disputes, and property management, tend to be disproportionately concerned by specific occupational groups (i.e., farmers, workers, and urban professionals, respectively). Moreover, the literature in sociolinguistics and social psychology has also demonstrated that socioeconomic status can be a critical factor that shapes individual’s sentence structure and word choices (e.g., Gumperz 1982; Labov 1964). While these patterns are not necessarily always obvious to human coders, some of them may be detected through an automated algorithms that systematically analyzes a large amount of text data.¹⁴

---

¹² Duplicates are identified based on a petition’s content and the time of posting (to the resolution of seconds).
¹³ Implemented in the R package ReadMe.
¹⁴ Our approach shares the spirit with research on computerized authorship attribution. For a review, see Stamatatos (2009).
To perform the estimation, we first take a random sample 200,000 petitions from the whole dataset and have a team of research assistants read through them and classify posters’ backgrounds into one of five mutually exclusive groups based on the petition content: (1) rural residents/farmers (农民), (2) workers (工人), (3) cadres and government employees (干部和编制内人员), (4) business people/vendors/managers (企业家/个体户/管理人员), and (5) other occupations. In cases of ambiguity, priority is given to the group of higher status. As a result, our coding approach is likely to underestimate the share of lower-class petitioners in the data.

Of the 200,000 petitions sampled, about 1/3 (~70,000) contains sufficient information for us to clearly identify the posters’ occupational background. We then use the 70,000 petitions as the training dataset to estimate the distribution of occupations in the entire body of petitions. Figure 3 shows the estimated proportions for these occupations and compare them with the same distribution among internet users in the nationally representative CCAS surveys. The results suggest that rural residents are the largest user group on the forum, accounting for about 40% of the petitions filed. This is especially remarkable given their relatively small share in the internet-using population. Another relatively over-represented group are workers, who account for only 7% of the online population but over 12% of the petitions in LLMB. By contrast, we find that only a small share of the petitions might have come from more privileged groups, such as business people and regime insiders (cadres and government employees). Overall, this pattern strongly supports our claim that the LLMB is indeed utilized disproportionately by people from humble social backgrounds.

To provide an intuitive summary of the issues raised by users, we further plot in Figure 4 the (translated) word clouds for both the entire sample of petitions and the five separate occupation groups. The size of each word is scaled to correspond to its frequency in the sample. Consistent with the content analysis, the full-sample word cloud demonstrates that rural residents and workers

---

15The classification is based on a codebook that lists a set of keywords for each of the groups. We provide a more detailed description of the coding procedures in the Online Appendix.
16For example, when a petition appears to be from a worker who also holds some managerial positions in a factory, it will be coded as from group 4 (business people/vendors/managers) instead of 2 (worker).
17Because the non-parametric method that we use for estimating occupation distribution does not generate classification for individual petitions, we only use the subset of manually coded petitions in plotting the occupation-specific word clouds.
are dominant user groups in this forum (as indicated by the large words for *villager*, *worker*, and *farmer*). It is also evident that most of the petitions are about mundane, personal issues rather than grand political demands. Words that most frequently used by rural petitioners appear to be related to issues such as land compensation, family planning, and access to basic services. For workers, their petitions appear to commonly touch on issues such as wage, pension, and labor contracts disputes.

6 The Policy Consequences of Online Participation

6.1 Data on the Number of Online Petitions

To analyze the policy consequences of online participation, we construct a panel dataset that records the volumes of LLMB petitions to all city-level governments in China between 2008 and 2013.\(^{18}\) We exclude three far-flung western provinces, Xinjiang, Tibet, and Qinghai. Given their geographic remoteness and distinct ethnic composition, we have reasons to believe that the pattern of participation and government policy making in those regions may be qualitatively different from others.\(^{19}\)

Our key independent variable of interest is the total number of petitions about a city government each year, which include both (A) petitions filed directly at the city’s own message boards and (B) petitions at the city’s supervising provincial government regarding issues in a specific city. For the second type of petition, we identify their city origins based on posters’ IP address and mentioning of city keywords in texts.\(^{20}\) Between 2008 and 2013, an average city receives about 153 petitions per year, with a large variation of 290. Figure 6 further illustrates the geographic variations in the average number of petitions per 10,000 residents during the same period. Interestingly, consistent

---

\(^{18}\)We focus on this period because it best matches with the availability of important socioeconomic covariates from other data sources

\(^{19}\)For example, during the period of interests, the average number of petitions filed in the three provinces is less than 1/10 of what were filed in the rest of the provinces.

\(^{20}\)When a petition mentions multiple city name keywords, we pick the unit that receives most mentioning as a petition’s origin city.
with the pattern at the individual level, the regional variations in aggregate participation also appear to be negatively associated with economic wealth: The more developed coastal areas actually saw lower per capita participation compared to many less developed, interior regions.

6.2 Data on Local Government Policy Priorities

We measure the policy priorities of local governments by analyzing the text of local Government Work Reports (GWRs 政府工作报告). GWRs are a form of official communication between the Chinese government and the legislative body at the same level. They are delivered once a year by the head of the government at the annual session of the People’s Congress (PC) and have to be formally approved by PC deputies through an anonymous vote. A GWR typically provides a comprehensive description of the policy blueprint that an administration has for the incoming year and highlights the key targets it seeks to achieve. As one of the most important policy documents issued by local governments each year, GWRs have to be collectively edited and approved by the party standing committee, which is the core leadership body in a regional local government, before they are sent out to the legislature for a vote.

GWRs have a highly standardized structure: They often begin with a short litany describing the overall national and local conditions, followed by a summary of the government’s achievements in the past year; the bulk of the document is then devoted to laying out plans and directions for the next year. While every GWR usually has to cover a fixed set of major policy areas (e.g., economy, public safety, culture, education, social welfare, etc), the amount of emphasis placed on each policy area is subject to political discretion and can reflect local leaders’ own policy visions. Newly published GWRs often receive close attention from media and government officials because they contain important information about leadership’s policy preferences.\(^{21}\)

We collect the full text of over 4,400 government reports at both the city and provincial levels between 2000 and 2014, and use a Latent Dirichlet Allocation (LDA) model (Blei, Ng and Jordan 2003) even financial investors would closely scrutinize these documents and search for new business opportunities that are in line with government policies. For an example, see http://www.jrzj.com/181254.html (in Chinese).
to uncover topics (i.e., cluster of words) from the text and estimate their relative proportions. Compared to the traditional, dictionary-based coding methods, a distinct strength of the LDA algorithm is that it clusters words strictly according to their co-occurrence patterns, thus avoiding the arbitrariness and errors in hand coding. This is especially valuable for analyzing complex policy documents such as the GWRs, for which commonly agreed coding rules simply do not yet exist. Moreover, since many words can be related to multiple policy areas, human coder may face difficulties in determining to which topic a particular word should be assigned, but the LDA model solves this problem by allowing a single word to be associated with multiple topics, albeit at different levels of strength.

We estimate a 10-topic model as the baseline but also a 20-topic model as a robustness check. Figure 5 plots the aggregate distribution of topic shares in the 10-topic model. In this model (as well as the 20-topic one), we are able to recover a clear and highly coherent topic cluster on social welfare issues, which on average accounts for about 8% of a GWR. The highest frequency words under this topic include social safety net (社会保障), pension coverage (养老保险), medical insurance (医疗保险), labor force (劳动力), social security (社会保险), and migrant workers (农民工). The following paragraph is taken from the 2009 GWR of Luoyang city, with key words under the social welfare topic highlighted in red.

[All levels and departments should]...carefully work on the issue of employment and reemployment, and fully implement supporting policies such as social security subsidies and tax breaks, with the goal of making 200 million yuan of microfinance loans in the whole year, adding 100 thousands urban jobs, and developing 1000 new commonweal posts...[We need to] encourage migrant workers to return home to set up businesses and to perfect the urban and rural social security system...[We also need to] strive to become a pilot city of rural pension, expedite the development of a social security system for landless farmers...[and] increase the construction of affordable

---

22 For example, a paragraph that contains the word infrastructure can be related to both the economic goal of promoting GDP growth and improving the living conditions of ordinary citizens.

23 In the interest of space, we leave the detailed output of the 10-topic model to the Online Appendix. The output from the 20-topic model is available upon request.
We perform several post-estimation diagnostics to evaluate the quality of the estimated topics. We focus on two metrics. The first one is coherence, which measures the tendency for top words in the topic to appear together; the second one is specificity, which indicates whether the distinctiveness of the words in a given topic compared to a uniform distribution of words (see the Online Appendix for more details). The social welfare topic is ranked as one of the best quality topics generated by both the 10- and 20-topic LDA models.\footnote{We provide more details about these diagnostic tests in the Online Appendix.} A variance decomposition exercise also suggests that within-city variation accounts for about 63% of the total variations in this topic, whereas between-city variations account for the rest of the 37%.

### 6.3 Empirical Strategy

We then proceed to estimate the effect of online participation on local governments’ emphasis on welfare-related policies. Given the panel structure of the data, the baseline is a fixed-effects model with the following specification:

\[
\text{Welfare Topic Share}_{it+1} = \delta \text{Log Petitions}_{it} + \mathbf{X}_{it}\boldsymbol{\beta} + \phi_i + \tau_t + \epsilon_{it},
\]

where \(i\) indexes cities and \(t\) indexes years. As discussed before, the main independent variable, \(\text{Log Petition}_{it}\) is a non-negative continuous variable for the (logged) number of city-specific petitions filed at city or provincial leaders’ message boards. The dependent variable, \(\text{Welfare Topic Share}\), is the estimated proportion of welfare topic in a city’s annual GWR. Because the delivery of the government work reports typically occurred at the beginning of a year (January or February), we match each city-year spell with GWR welfare topic in the next year. \(\phi_i\) and \(\tau_t\) represent city and year fixed-effects, respectively. The main regression also includes a set of province-specific linear trends to account for influence from time-variant, province-specific unobserved factors.\footnote{An alternative specification is to use province-year dummies, which would consume more degrees of freedom in the estimation. Our main result is substantively the same under the alternative specification.}
Furthermore, $X_t$ encompasses a rich set of time-variant covariates that can be important potential confounders to the main quantity of interest. We detail these covariates below.

**Frequency of Offline Protests**

The most important potential confounder to our analysis is the intensity of public grievances expressed through other non-internet channels. Since dissatisfied citizens may concurrently use multiple means to make their grievances heard, changes in government policies may be attributable to changes in their offline actions rather than an increase in online participation per se. According to prior research, one of the most common alternative strategies that lower-class citizens in China use to express their discontent is through collective protests (Chen 2013; O’Brien and Li 2006); anecdotes suggest that large-scale protests tend to receive tremendous attention from the political authority and sometimes produce major shifts in government policies.\(^{26}\)

While the government does not provide official statistics on the frequency of protests, several research teams both within and outside China have attempted to systematically collect such data using different methods. Here, we make use of the two of the largest available datasets. The first one is the Collective Incidents Dataset, compiled by the institute of sociology at the Chinese Academy of Social Sciences (CASS).\(^{27}\) The CASS dataset contains detailed information about major mass protests in China between 2007 and 2013, collected from both internal government documents and through extensive search of domestic and overseas media reports. The second one, *China Strikes* (www.chinastrikes.crowdmap.com), is a crowd-sourced website that focuses specifically on labor-related unrest.\(^{28}\) For our analysis below, we combine all the unique records from both sources into a new dataset, which contains a total of 1268 protests between 2008 and 2013. On average, a city experiences about 0.7 protests per year with a large standard deviation of 2.4. A

\(^{26}\)For an example, see Lagerkvist (2012).

\(^{27}\)The dataset can be accessed at http://www.cssn.cn/sjxz/yhfw/qtxjsjk/qtxsjkj/201312/t20131211_903667.shtml.

\(^{28}\)Even though these two datasets are supposedly collected by different teams and through different methods, we find a correlation of 0.74 between the two in terms of aggregate number of protests at the city level. This large correlation gives us confidence that there is a common underlying pattern of social instability that both datasets have managed to capture.
simple correlation test suggests there is indeed a positive and significant association between the frequency of offline protest and the volume of online petition ($\rho = 0.23, p < 0.0001$).

**Socioeconomic Conditions**

The second set of covariates that we include are intended to important local socioeconomic conditions. We include controls for a city’s total population ($\text{Log Population}$) and the size of employment ($\text{Log Employment}$) as unemployment is a common source of popular grievances and a main driver for welfare expansion. Moreover, because welfare-related spending in China is covered primarily by local government budgets, the breadth and depth of welfare provision is often a function of local economic and fiscal conditions (Huang 2015). We thus also include variables such as $\text{Log GDP}$, $\text{GDP Growth Rate}$, $\text{Log Fiscal Revenue}$, and $\text{Log Fiscal Expenditure}$ to control for a locality’s level and pace of development and its government’s fiscal capacity.

**Leadership Characteristics**

The third set of controls concern personal characteristics of local leaders. We include a number of demographic variables (for both the city secretary and the mayor), such as age, gender, education, tenure length, and political connection with the provincial leadership. Age and tenure length, in particular, have been found to be highly correlated with the career incentives of local leaders as well as their policy preferences (e.g., Guo 2009). We also include the total length of time they have served in a given city ($\text{City Secretary/ Mayor’s Years of Local Service}$), as prior research suggests that officials with longer local careers may be more attentive to local interests and spend more on social issues (Persson and Zhuravskaya Forthcoming).

**6.4 Baseline Results**

The main results are presented in Table 1. We begin with the most parsimonious model with only two sets of fixed-effects and linear province trends as controls, and then incrementally add the socioeconomic and city leader controls. The key coefficient estimate remains highly stable as
more covariates are included. Overall, the results suggest that the frequency of online petitions is strongly and positively associated with the proportion of the welfare topic in GWRs. Focusing on Column 3, the result suggests that for an average city, a one standard deviation increase in online petition is associated with a 0.34 percentage point increase in the welfare topic share. Since the average share of welfare topic is about 8 percent in our sample, this magnitude amounts to about 4.3 percent increase from the baseline proportion. In Columns 4 to 6, we replicate the same analysis using a dependent variable generated from a 20-topic model, and find essentially identical results.

We conduct several robustness checks of the main results but will leave the details to the Online Appendix in the interest of space. To briefly summarize, we find that our main results are robust to (1) using incremental changes in (as opposed to the levels of) welfare topic as the dependent variable and (2) alternative estimation strategies (beta regression and fractional logit) that takes into account the proportional nature of the dependent variable, and (3) a permutation test that randomly reshuffles the correspondence between city identities and petition volumes. Furthermore, a placebo test that manipulates the timing of the dependent variable suggests that the volume of online petitions at t is only significantly associated with GWR welfare topic in the next year (t + 1), but not with that in earlier years (t or t − 1).

### 6.5 Effects of Online Participation on Substantive Outcomes

Next, we examine other substantive policy consequences associated with online participation in addition to changes in GWR topics. We begin by investigating whether online participation has produced meaningful changes to local governments’ policy agenda by looking at a much larger set of official documents and legislation that they produce. To do so, we make use of PKU-LAW ([www.pkulaw.cn](http://www.pkulaw.cn)), a comprehensive database on Chinese laws and regulations maintained by Peking University. We conduct a series of keyword searches in the database using the top 30 words under the GWR welfare topic, and obtain the total count of welfare-related legislation for each city-year spell. Column 2 of Table 2 presents results from a regression using the count number
as the dependent variable. The result suggests that the scale of online participation also positively and significantly affects the level of policy/legislative activities in the area of social welfare policies. For a one standard deviation increase in online participation, the number of welfare-related laws and regulations issued will on average increase by about 10% of a standard deviation in the following year.

Another way to assess the substantive impacts of online participation is to look at changes in government spending patterns. We collect data on three welfare-related areas in which fiscal statistics are available—social security (shehui baozhang), medicine and health (yiliao weisheng), and education (jiaoyu), and regress their annual log changes on our participation variable. The results, displayed in Columns 2 through 4 in Table 2, suggest that increases in online petition is associated with significantly more rapid growth in social security spending and positive, but insignificant, changes in the other two areas. To further check whether the growth in welfare spending is driven by a general increase in fiscal expenditure, we also run a regression using log change in non-welfare expenditure the dependent variable and report the results in the last column of Table 2. As the expected, we find that online participation has little bearing on growth in fiscal outlays that are unrelated to social welfare.

6.6 Effects of Participation by Petition Topics

We also explore which types of petitions are most closely associated with changes in welfare policies. To do so, we first run a 30-topic LDA model on the petition data and then label each petition by the topic that takes up the highest proportion in its text. For each topic \( T (1 \leq T \leq 30) \), we estimate the following regression

\[^{29}\text{Non-welfare expenditure is calculated by subtracting from total expenditure the spending in the three aforementioned areas.}\]
Welfare Topic Share$_{it+1} = \delta^T \log\text{Petitions in Topic } T_{it}
+ \delta^{NT} \log\text{Petitions Not in Topic } T_{it}
+ X_{it}\beta + \phi_i + \tau_t + \epsilon_{it}.$

We plot the four most positive and most negative $\delta^T$’s (a total of 8 coefficients) in Figure 7 along with their 90% confidence intervals (this includes all the estimates that are statistically significant). We also print the ten keywords that have the highest association with each topic on top of its coefficient estimate. Reading through the keywords confirms that most of the topics that have the greatest positive association with changes in welfare topics are about livelihood issues cared primarily by rural, lower-class citizens: Topic 16, for example, is about land disputes and compensation, which mostly concern farmers and rural residents. Topic 7 encompasses several issues related to migrant workers, such as local resident permit, household registration, and birth planning. Topic 21 also includes a number of words that directly refer to disadvantaged groups, such as low-income households and individuals with disabilities. The only exception, however, seems to be Topic 10, which on surface appears to be a topic about transportation and vehicle registration. After taking a closer look at the context of these petitions, however, we find out that a large number of complaints under this topic actually came from farmers who had trouble registering/reporting accidents about their trucks, tractors or other agriculture vehicles in remote villages.

Next, we turn to those topics that have the most negative loadings on the welfare topic share. According to the keywords displayed, both Topics 18 and 4 appear to be about general appellations and salutations. Topics 22 is one about problems in water supply, TV signals, and electricity, and Topic 8 about transportation and tourism. Overall, these topics seem to have much less direct relevance to social welfare issues than those with positive loadings. The qualitative contrast between these two sets of topics thus gives us with greater confidence that there is a substantive correspondence between issues articulated in the online participation channel and subsequent government
policy changes.

6.7 Evaluating Potential Mechanisms

The preceding analyses have shown that increased mass participation via the LLMB has resulted in a notable shift in government policies toward redistributive and social welfare issues. A key remaining question, however, is what induced unelected local authorities to make policy responses to these online petitions? This section evaluates three plausible mechanisms: (1) information revelation, (2) amplifying threats of collective actions, and (3) facilitating top-down monitoring.

6.7.1 Information Revelation

The first possible mechanism is that a large inflow of petitions from previously underrepresented groups might have simply primed local governments those groups’ presence and made the authority more aware of their demands (Iyengar, Peters and Kinder 1982). This mechanism rests on the assumption that local policymakers possess a general inclination to serve the interests of their constituencies, but their ability to make the right policies are somewhat constrained by the quality and representativeness of the information available to them. Although city leaders are (obviously) not electorally accountable to their constituencies in China, they may feel a normative pressure to be responsive to local citizens as both the Confucian political culture and the Communist mass-line ideology place heavy emphasis on political leaders’ moral obligation to faithfully serve the people (Tang 2016; Yang and Zhao 2015). Anecdotes also suggest that many officials care deeply about their personal reputations among local citizens (koubei). It is therefore possible that exposure to large volumes of grievances from underprivileged citizens has increased the relative salience of social welfare issues in the minds of the local officials, leading them to place greater weight on these issues when crafting policies.

We assess the validity of this mechanism by empirically testing some of its key implications. According to a large psychology literature, the effectiveness of priming depends critically on the

30For an example, see http://news.cctv.com/society/20081203/102490.shtml.
availability of the primed information at a time when relevant decisions are made (e.g., Bargh and Chartrand 2014). In the context of our study, this implies that certain subsets of the petition may be better at influencing government policies because they are filed in the “right” places or moments that are more likely to be noticed and memorized by the authority. We investigate this possibility in several ways. First, we compare the policy effects of petitions filed at three different levels—(1) the supervising provincial government, (2) the city government, or (3) the county-level governments within a city. While a citizen living in a city can in theory register his/her complaints to any of the three levels, those sent to the city’s own message boards can usually be most easily observed and noted by city leaders. Therefore, if the priming mechanism is true, we should expect that the scale of petition at a city’s own message boards to be more closely associated with policy changes than those at its subordinate or supervising governments’.

Columns 1 through 4 of Table 3 present results that confirm this conjecture. Here, we first run three separate regressions of welfare topic proportion on each type of petitions and then a forth regression that includes all three types. We find that although the volumes of all three types of petitions track closely with each other, only city-level petitions seem to have a strong policy impact. By contrast, the coefficient estimates for petitions at the provincial and county-level, while positive, are much smaller in size and statistically insignificant in general.

Relatedly, information availability may also vary along the temporal dimension. For example, since the preparation of the GWR usually begins during the second half of the previous year, according to the well-known literature on recency bias (e.g., Weingast, Shepsle and Johnsen 1981), we should expect petitions filed during the preparation months to have greater influence on the content of the GWR than those filed during earlier periods. Similarly, because government agents who are responsible for collecting and reporting online petitions typically only work on weekdays, we should also expect petitions filed between Monday and Friday to have a larger influence on policies than those filed over the weekends.

These hypotheses are tested in the last two columns of Table 3. Column 5 reports results from

31 Provincial vs. city: \( \rho = 0.80 \); provincial vs. county: \( \rho = 0.64 \); city vs. county: \( \rho = 0.67 \).
a regression that includes petition counts for three, four-month intervals—January to April, May to August, and September to December, respectively. We find that the coefficient estimate is largest and only significant for petitions filed between August and December. Column 6 further shows that, compared to petitions filed over weekends, those filed during the weekdays tend to be more closely associated with policy changes.\textsuperscript{32} Taken together, these results provide consistent evidence that information revelation is indeed an important mechanism that translates online participation into policy changes.

### 6.7.2 Amplifying Threats of Collective Actions

A second plausible mechanism is that online participation amplified the threats of collective actions by lower-class citizens. It is widely documented that the Chinese authority is preoccupied with maintaining social stability and has devoted tremendous resources into preventing and preempting collective actions (Chen 2013; King, Pan and Roberts 2013). A large volume of petitions can be seen as a precursor to more significant offline protests; and the transparent nature of the forum can be especially conducive for citizens with common grievances to find each other and organize. Local governments facing large volume of online petitions may thus be motivated to make policy concessions as a way to preempt potentially more disruptive off-line collective actions.\textsuperscript{33}

Table 4 presents several regressions that test various aspects of this mechanism. Column 1 reports estimates from a regression that uses reply rate to individual petitions as the dependent variable. The results indicate that the frequency of protests is positively associated with reply rates to individual petitioners. This finding is also consistent with those from recent studies such as Chen, Pan and Xu (2016) and Distelhorst and Hou (Forthcoming), both of which have shown that the threat of collective action can significantly increase the reply rates to individual petitions on government websites.

\textsuperscript{32}It is also worth noting that the estimates become much noisier when we include multiple groups of petitions at the same time. This is in part due to the fact that petition volumes in these groups are highly correlated.

\textsuperscript{33}As a piece of evidence on the collective-action-preventing motives, many localities have reported a dramatic drop of offline petition by local residents after the online channel became available (Wang 2010), suggesting that these two methods are often substitutes.
In Column 2, we further estimate a regression that includes an interaction term between online petition and the cumulative number of protests that a city has experienced since 2007\textsuperscript{34}. The cumulative measure is intended to capture city residents’ long-term potential capacity (and propensity) to engage in collective actions. If the collective action threat mechanism holds, it is natural to expect that local governments’ responsiveness to online participation would be higher in cities that have experienced more instability in the past. Consistent with this expectation, we find that the coefficient estimate for the interaction term is positive and statistically significant. The magnitude of the coefficient suggests a sizable substantive effect: For a one standard deviation in the cumulative protest intensity (~7.83), the marginal impact of online petition on welfare topic share will increases by over 60% (from 0.28 to 0.47 percentage points).

\subsection*{6.7.3 Facilitating Top-down Monitoring}

Finally, the third mechanism we evaluate is that online participation drives policy changes by improving top-down monitoring within the government. Since the forum provides direct and largely uncensored information about citizen demands, it might have also served as a useful avenue for senior leaders (at provincial level or higher) to learn about local grievances and the behaviors of their subordinates. Anecdotes suggest that some provincial leaders were indeed actively using the forum to reach out to local citizens and to make “incognito inspections” over local cadres (Wang 2010); this could in turn give local leaders an incentive to act in a responsive fashion as a way to to impress their superiors or simply to reduce the negative publicity of citizen complaints.

To evaluate this possibility, we interact the main petition variable with several other variables that measure provincial leaders’ interest in the LLMB as well as the cost of monitoring through conventional bureaucratic methods. We use both reply rates and total number of petitions in provincial leaders’ own message boards to capture provincial leader’s engagement with the petition forum. We also calculate the log distance between a city and its provincial capital as a proxy for monitoring cost. The results, which are shown in Table 5, indicate that magnitudes of coefficients

\textsuperscript{34}2007 is the starting year of our protest dataset.
for all three interactions are rather small and none of them reach the conventional level of statistical significance. While we cannot completely rule this mechanism simply based on these null findings, it does seem that the empirical evidence does not provide as strong support for it as for the other two.

7 Conclusion

Do new communication technologies mitigate or exacerbate inequalities in political influence among citizens? This study presents new evidence to this debate by analyzing how the emergence of an internet-based petition forum affects citizen participation and government policy making in China, a non-democratic country with the world’s largest number of internet users. We show that the forum is disproportionately used by citizens from lower-class backgrounds, and that online participation via this forum has led local governments to place greater emphasis on social welfare issues in policy rhetoric, legislation, and fiscal spending. Furthermore, we provide evidence that helps adjudicate three potential mechanisms that link online participation to policy change. Our findings suggest that online participation affects local leaders’ policy decisions by both providing information about the needs of traditionally underrepresented groups and by amplifying the threats of more disruptive, offline collective actions. Meanwhile, we do not find evidence that better monitoring from direct political superiors is a main mechanism that drives policy changes.

The finding that the internet can help mitigate de facto political inequality among citizens carries important practical implications. Research has shown that a high level of inequality tends to breed numerous socioeconomic problems such as crimes, conflicts and under-provision of public goods (Thorbecke and Charumilind 2002), and that the root of social and economic inequalities often lies in the political realm (Acemoglu et al. 2007). The advent of new technology thus may present a unique opportunity to address these intertwined problems by giving disadvantaged citizens greater influence over the policy making process. While we do need to be cautious about over-generalizing our findings about this particular forum to other settings, it is also worth noting
that some of the key features that made the forum effective, including operational independence, ease of access, and transparency, are not rare qualities in the cyberspace. We therefore have reasons to expect similar equalizing effects when this model is replicated in other political systems.

By showing that a nondemocratic government not only tolerates, but also encourages and actively responds to, opinions and demands made on the internet, this study offers a more nuanced perspective on the relationship between the internet and authoritarian regimes. While previous studies tend to portray a largely confrontational picture, focusing on either the internet's capacity to subvert authoritarian rule (Ferdinand 2000) or the regime's efforts to control and manipulate the online sphere (Kalathil and Boas 2010; King, Pan and Roberts 2013), our analysis suggests that the interaction between the two is not always zero-sum. Some authoritarian regimes may be able to harness the democratic potential of the internet and use it to fulfill important governing functions, such as gathering information, channeling social contention, and strengthening bottom-up accountability. These benefits may, paradoxically, help prolong the rule of those regimes by allowing them to improve the quality of governance without making fundamental changes to their political systems.
References


30


Smith, Aaron. 2014. “African Americans and Technology Use.”.

URL: [http://www.pewinternet.org/2014/01/06/african-americans-and-technology-use/](http://www.pewinternet.org/2014/01/06/african-americans-and-technology-use/)


Wang, Donglian. 2010. “Wangluo Wenzheng yu Zhengfu Shanzhi (Internet-based Political Participation and Good Governance).”  
**URL:** [http://theory.people.com.cn/GB/11936154.html](http://theory.people.com.cn/GB/11936154.html)


Figures and Tables

Figure 1: Online Participation is Less Sensitive to Income

Note: This figure illustrates the relationship between participation modes and family income using the 2008 and 2009 Chinese Citizenship Awareness Surveys. The blue solid line indicates the relationship between income and traditional mode of participation, which include (1) attending government meeting, (2) making appeals to the upper-level government, and (3) contacting media, local cadres, or People’s Congress deputies to express their views. Both the green dotted line and the red dashed line illustrate the relationship between income and participation via online channels, based on the sample of all internet users and frequent internet users (weekly usage), respectively.
Figure 2: A Snapshot of the LLMB

(a) Entry Interface

(b) Message-viewing Interface

Note: This figure displays a snapshot of the actual interface of the Local Leader Message Board as of May 15, 2017. Key functions and message contents are translated by the authors.
Figure 3: Occupational Distribution Compared: LLMB User vs. Nationally Representative Surveys

Note: This figure compares the estimated occupation distribution from the LLMB (red bar) with the occupation distribution of internet users in the nationally representative CCAS surveys (blue bar).
Figure 4: Word Clouds by Occupation

(a) Full-Sample

(b) Rural Residents

(c) Workers

(d) Business Owners

(e) Cadres and Government Employees

(f) Other Occupations
Figure 5: Aggregated Distribution of GWR Topic Shares, 2000-2014
Figure 6: Average Number of Petitions per Capita, 2008-2013

- Petitions per 10,000 residents
- No data/Not a prefecture unit
- [0.05,0.13) (<20%)
- [0.13,0.19) (20−40%)
- [0.19,0.28) (40−60%)
- [0.28,0.51) (60−80%)
- [0.51,3.96] (>80%)
Figure 7: Effect of Participation by Petition Topic

Note: This figure illustrates the coefficient estimates for the effect of the volume of topic-specific petitions on welfare topic proportions. The circles indicate the point estimates and the horizontal bars the 90% confidence intervals. The 10 words with the highest frequency under each topic are printed.
Table 1: Main Results

<table>
<thead>
<tr>
<th></th>
<th>Welfare Topic at $t + 1$ (10-topic)</th>
<th>Welfare Topic at $t + 1$ (20-topic)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Log petitions</td>
<td>0.0033**</td>
<td>0.0034**</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>Log protests</td>
<td>-0.0022</td>
<td>-0.0019</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0015)</td>
</tr>
<tr>
<td>Log employment (10,000 persons)</td>
<td>0.0006</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>(0.0036)</td>
<td>(0.0034)</td>
</tr>
<tr>
<td>Log GDP</td>
<td>-0.0035</td>
<td>-0.0037</td>
</tr>
<tr>
<td></td>
<td>(0.0091)</td>
<td>(0.0100)</td>
</tr>
<tr>
<td>Log population</td>
<td>-0.0113</td>
<td>-0.0085</td>
</tr>
<tr>
<td></td>
<td>(0.0069)</td>
<td>(0.0071)</td>
</tr>
<tr>
<td>GDP growth rate</td>
<td>-0.0005**</td>
<td>-0.0005**</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Log fiscal expenditure</td>
<td>0.0050</td>
<td>0.0055</td>
</tr>
<tr>
<td></td>
<td>(0.0091)</td>
<td>(0.0089)</td>
</tr>
<tr>
<td>Log fiscal revenue</td>
<td>-0.0005</td>
<td>-0.0013</td>
</tr>
<tr>
<td></td>
<td>(0.0055)</td>
<td>(0.0056)</td>
</tr>
<tr>
<td>City secretary’s age</td>
<td>0.0004</td>
<td>0.0004*</td>
</tr>
<tr>
<td></td>
<td>(0.0002)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>Mayor’s age</td>
<td>0.0002</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>City secretary’s tenure</td>
<td>0.0003</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>Mayor’s tenure</td>
<td>0.0006</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>City secretary’s years of local service</td>
<td>-0.0003**</td>
<td>-0.0002*</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Mayor’s years of local service</td>
<td>0.0002</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Connected city leader</td>
<td>0.0005</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>(0.0017)</td>
<td>(0.0016)</td>
</tr>
</tbody>
</table>

City and year fixed-effects ✓ ✓ ✓ ✓ ✓ ✓
Province-specific trends ✓ ✓ ✓ ✓ ✓ ✓
Adjusted $R^2$ 0.25 0.25 0.26 0.26 0.26 0.26
Number of Cities 300 300 300 300 300 300
Observations 1719 1687 1685 1719 1687 1685

Note: This table shows the results from OLS regressions. The dependent variables are incremental increase in the share of social welfare topic in government work reports (for both the 10-topic model and the 20-topic model). Robust standard errors clustered at city level are reported in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (two-tailed test)
Table 2: Change in Substantive Outcomes

<table>
<thead>
<tr>
<th>Welfare legislation at ( (t + 1) )</th>
<th>( \Delta ) Log Fiscal expenditure</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log petitions</td>
<td></td>
<td>0.0908***</td>
<td>0.0473**</td>
<td>0.0049</td>
<td>0.0238</td>
<td>-0.0017</td>
</tr>
<tr>
<td>City and year fixed-effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Province-specific trends</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Adjusted R(^2)</td>
<td>0.41</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>Number of Cities</td>
<td>295</td>
<td>304</td>
<td>304</td>
<td>304</td>
<td>304</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1689</td>
<td>1810</td>
<td>1814</td>
<td>1799</td>
<td>1795</td>
<td></td>
</tr>
</tbody>
</table>

Note: The table reports results from regressing other policy outcomes on petition volumes. The specification is based on Column 3 of Table 1. Column 1 additionally controls for the total (logged) number of laws and regulations passed in a year. Robust standard errors clustered at city level are reported in parentheses. * \( p < 0.1 \), ** \( p < 0.05 \), *** \( p < 0.01 \) (two-tailed test)

Table 3: Mechanism 1: Information Revelation

<table>
<thead>
<tr>
<th>Welfare Topic (10-topic)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log petition (provincial-level)</td>
<td>0.0012</td>
<td>0.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log petition (city-level)</td>
<td>0.0027***</td>
<td>0.0026***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td>(0.0010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log petition (county-level)</td>
<td></td>
<td>0.0008</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0009)</td>
<td>(0.0010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log petition (Sep-Dec)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0025*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0013)</td>
</tr>
<tr>
<td>Log petition (May to Aug)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0015)</td>
</tr>
<tr>
<td>Log petition (Jan to Apr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0014)</td>
</tr>
<tr>
<td>Log petition (weekdays)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0036*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0020)</td>
</tr>
<tr>
<td>Log petition (weekends)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0017)</td>
</tr>
</tbody>
</table>

Note: The table reports results from testing several implications following the priming mechanism. The specification is based on Column 3 of Table 1. Robust standard errors clustered at city level are reported in parentheses. * \( p < 0.1 \), ** \( p < 0.05 \), *** \( p < 0.01 \) (two-tailed test)
Table 4: Mechanism 2: Amplifying the Threats of Collective Actions

<table>
<thead>
<tr>
<th></th>
<th>Reply Rate</th>
<th>Welfare Topic at t + 1 (10-topic)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Log petitions</td>
<td>0.0703***</td>
<td>0.0029**</td>
</tr>
<tr>
<td></td>
<td>(0.0185)</td>
<td>(0.0015)</td>
</tr>
<tr>
<td>Log protests</td>
<td>0.0253**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0103)</td>
<td></td>
</tr>
<tr>
<td>Log petition × Cumulative # of protests (since 2007)</td>
<td>0.0002***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td></td>
</tr>
</tbody>
</table>

City and year fixed-effects ✓ ✓
Province-specific trends ✓ ✓ ✓
Adjusted R² 0.21 0.26
Number of Cities 303 300
Observations 1778 1685

Note: The table reports results from testing several implications following the collective action threat mechanism. The specification is based on Column 3 of Table 1. All the individual terms of the interactions are included in regressions. Robust standard errors clustered at city level are reported in parentheses. ∗ p < 0.1, ∗∗ p < 0.05, ∗∗∗ p < 0.01 (two-tailed test)

Table 5: Mechanism 3: Top-Down Monitoring

<table>
<thead>
<tr>
<th></th>
<th>Welfare Topic (10-topic)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3)</td>
</tr>
<tr>
<td>Log petitions × Log total petition to provincial leaders</td>
<td>-0.0000 (0.0008)</td>
</tr>
<tr>
<td>Log petitions × Reply rate by provincial leaders</td>
<td>-0.0013 (0.0030)</td>
</tr>
<tr>
<td>Log petitions × Distance to provincial capital</td>
<td>-0.0007 (0.0006)</td>
</tr>
</tbody>
</table>

City and year fixed-effects ✓ ✓ ✓
Province-specific trends ✓ ✓ ✓
Adjusted R² 0.26 0.26 0.26
Number of Cities 300 300 300
Observations 1685 1685 1685

Note: The table reports results from testing several implications following the monitoring mechanism. The specification is based on Column 3 of Table 1. All the individual terms of the interactions are included in regressions. Robust standard errors clustered at city level are reported in parentheses. ∗ p < 0.1, ∗∗ p < 0.05, ∗∗∗ p < 0.01 (two-tailed test)
Online Appendix for

“From Internet to Social Safety Net: The Policy Consequences of Online Participation in China”

Table of Contents

A  Divergent Preferences between the Rich and the Poor               A-2

B  Details on Estimating Government Policy Priorities                A-3
   B.1 Methodology                                                  A-3
   B.2 Estimation Procedure                                          A-4
   B.3 Post-Estimation Diagnostics                                   A-7
   B.4 Estimated Government Work Report Topics                      A-9

C  Distribution of Logged Number of Petitions by Year               A-13

D  Data on Collective Protests                                     A-13

E  Using Semi-automated Content Analysis to Identify Backgrounds of Petitioners  A-14

F  Robustness Checks                                               A-17
   F.1 Alternative Estimation Strategy                              A-17
   F.2 Dealing with Reverse Causality                              A-18
   F.3 Effect of Participation over Time                           A-20
   F.4 Permutation Inference                                        A-21

G  Mapping Responsiveness by Province                               A-22
A Divergent Preferences between the Rich and the Poor

This section draws on data from the 2008 and 2009 China Citizenship Attitude Surveys (CCAS) to illustrate the diverging preference between the poor and the rich in China. Specifically, we utilize a questionnaire item (D10) that asks the respondents to pick an issue that they think are the “most important” from the following 4 options:

1. Maintaining domestic order
2. Give people more voice in government decision-making
3. Protect civil liberty
4. Improve welfare of the poor

We estimate a multinomial logit model where individual choices are regressed on log family income along with other covariates. Table A.1 presents the results, which are transformed into marginal effects for the ease of interpretation. The results suggest that family income is a highly significant predictor of preferred policy priorities. Compared to wealthy, lower-income respondents are much more likely to regard welfare improvement for the poor as the most important issue and much less likely to place priority on either public order or policy voices. Consistent with the pattern, we also find that the probability of choosing “don’t know” decreases as income rises. These results thus support our claim that the rich and the poor possess distinct preferences for policy issues in contemporary China.
Table A.1: Different Demands from High- and Low-Income Individuals

<table>
<thead>
<tr>
<th>DV: Most Important Task for the Government</th>
<th>(1) Public order</th>
<th>(2) Voice on policy</th>
<th>(3) Freedom</th>
<th>(4) Improve welfare</th>
<th>(5) Don’t Know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log family income</td>
<td>0.026***</td>
<td>0.013***</td>
<td>0.005</td>
<td>-0.024***</td>
<td>-0.020***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.009)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Age</td>
<td>0.000</td>
<td>0.000***</td>
<td>-0.000**</td>
<td>0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.086***</td>
<td>-0.008</td>
<td>-0.015</td>
<td>0.055***</td>
<td>0.053***</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.007)</td>
<td>(0.010)</td>
<td>(0.016)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>College education</td>
<td>0.262***</td>
<td>0.037</td>
<td>0.083**</td>
<td>0.456***</td>
<td>-0.838***</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.023)</td>
<td>(0.035)</td>
<td>(0.093)</td>
<td>(0.134)</td>
</tr>
<tr>
<td>CCP membership (self or family)</td>
<td>0.050***</td>
<td>0.002</td>
<td>0.006</td>
<td>-0.031</td>
<td>-0.027**</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.009)</td>
<td>(0.012)</td>
<td>(0.021)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Employed in state sector</td>
<td>0.023</td>
<td>-0.003</td>
<td>-0.002</td>
<td>0.007</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.012)</td>
<td>(0.016)</td>
<td>(0.030)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Rural residency</td>
<td>-0.067***</td>
<td>-0.011</td>
<td>-0.013</td>
<td>0.039</td>
<td>0.053***</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.010)</td>
<td>(0.014)</td>
<td>(0.025)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Province dummy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Year dummy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Baseline probability</td>
<td>0.305</td>
<td>0.052</td>
<td>0.102</td>
<td>0.475</td>
<td>0.066</td>
</tr>
<tr>
<td>Observations</td>
<td>4142</td>
<td>4142</td>
<td>4142</td>
<td>4142</td>
<td>4142</td>
</tr>
</tbody>
</table>

Note: This table shows results from a multinomial logit model where the dependent variable is the respondent’s choice of the "most important task for government to accomplish". The result suggests that family income is negatively associated with the tendency to select welfare improvement as the most important task for the government. In other words, poor individuals care much more about social welfare issues than wealthy ones.

Data: China Citizenship Attitude Survey (2008 & 2009)
Robust standard errors clustered at individual level are reported in parentheses.
* p < 0.1, ** p < 0.05, *** p < 0.01 (two-tailed test)

B Details on Estimating Government Policy Priorities

B.1 Methodology

We use a Latent Dirichlet Allocation (LDA) model (Blei, Ng and Jordan 2003) to estimate the government’s attention to different policy areas in the work reports. As we will show, the algorithm discovers highly substantive and coherent topics, and measurement errors in estimated topic proportions are likely to be small. The advantages of estimating topic models over human reading and coding in our case are threefold. First, the algorithm clusters words strictly according to their
co-occurrence patterns, thus avoiding the arbitrariness and errors in hand coding. The somehow ambiguous boundary between topics and the fluid nature of language makes it challenging for a human coder to consistently parse the text with some pre-specified rules, where such rules may be hard to define in the first place. An LDA can not only group together words with similar semantics, but also words conforming to similar wording styles. This later type of distinction is particularly elusive to human eyes. Second, many policies are multi-faceted, and can be attributed to more than one areas. For example, a discussion of building infrastructure is related to both the economic goal of promoting GDP growth and the welfare goal of improving people’s living conditions. In this situation it is not immediately clear how one should code this piece of discussion. LDA solves this issue by allowing both topics to give rise to the word Infrastructure, albeit possibly with different probabilities. Which topic a particular occurrence of Infrastructure belongs to is then obtained via Bayesian estimation. Last but not least, using an automated algorithm tremendously reduces the cost of parsing thousands of lengthy and dry policy documents, making the documents a new source of data, which is unattractive, if not impossible, to utilize without the algorithm.

One might not want to take a topic model literally. That is, the data generating process of a topic model could be very different from the way these reports are actually written. However, LDA, the most basic form of probabilistic topic models, has been shown to exhibit very good performance in a wide range of applications. Blei (2012) provides a survey of the fruitful applications of LDA in political science, psychology, population genetics, computer vision, etc. Our corpus of government work reports turns out to be a particularly good testing field for LDA, because the formal and precise wording in the reports greatly reduces noise. As a result, the topics are well-demarcated, and the majority of them are substantive (policy relevant) topics instead of “wording style” clusters.

B.2 Estimation Procedure

We first do word segmentation on the government work reports. Chinese characters are not naturally separated from each other as in many Indo-European languages. Therefore one has to start with segmenting and demarcating the text. We apply the Jieba segmentation module (qinwf 2016)
to the corpus, which uses a maximum probability segmentation model and a Hidden Markov Model (HMM) to do the segmentation. The algorithm combines an existing dictionary of Chinese words and the ability to learn new words from the text. The Jieba package has a proven record in word segmentation, and is 5-20 times faster than other packages. We examined the segmented text, and confirmed that the segmentation quality is very good. Special terms such as Three Represents, Deng Xiaoping Theory and 18th National Congress of the Communist Party of China are correctly identified as a single word. The segmented text is naturally tokenized, and each piece becomes a token in subsequent computation. An unintended advantage of word segmentation is that our tokens are meaningful phrases instead of simple unigrams. Typical tokens are like reform and open up, urban and rural residents and poverty alleviation and development. These phrases clearly contain more information than unigrams.

We remove all punctuation, numbers and English words (such as GDP) from the text to focus on the Chinese vocabulary. We also remove a standard list of Chinese stop words - common words that are not really meaningful. These include words such as some, both and why. Because there is no inflection in Chinese, one does not need to stem the text (i.e., reducing words to their root forms). Each word in our vocabulary denotes a unique meaning. The precise and informative nature of the government reports further adds to the information contained in the preprocessed text.

Due to properties of the Dirichlet distribution, the algorithm tends to spread a topic across few words and a document across few topics. Therefore LDA is able to find topics of much higher quality when we define each paragraph of the reports as a document w, instead of defining each report as a document, as a paragraph is more likely to focus on a single issue. We train LDA on reports spanning from the year 2000 to the year 2013. These reports contain 440,202 paragraphs and 3,426,528 tokens in total. The LDA is thus trained on 440,202 documents. Once the model is trained with paragraphs as documents, we infer topic proportions in whole reports with Gibbs-sampling style re-sampling.

The data generating process of LDA is as follows.
A document \( \mathbf{w} = (w_1, \ldots, w_N) \) contains \( N \) words. Each word \( w_i \) is a member of the vocabulary of \( V \) words \( \{1, \ldots, V\} \). There are \( k \) topics, \( (z_1, \ldots, z_k) \). \( k \) has to be specified by the modeler. Each topic is a distribution over the words. These are characterized by a \( k \times V \) matrix \( \beta \), where \( \beta_{ij} = p(w^j = 1 | z^i = 1) \).

**Step 1:** The term distribution \( \beta \) is drawn for each topic from a Dirichlet distribution with parameter \( \delta \)

\[
\beta \sim \text{Dirichlet}(\delta)
\]

**Step 2:** The topic proportions of document \( \mathbf{w} \) are drawn from a Dirichlet distribution with parameter \( \alpha \)

\[
\theta \sim \text{Dirichlet}(\alpha),
\]

so that \( \theta = (\theta_1, \ldots, \theta_k) \), where \( \theta_i \) is the proportion of topic \( i \).

**Step 3:** For each of the \( N \) words \( w_n \),

(a) Choose a topic \( z_n \sim \text{multinomial}(\theta) \).

(b) Choose a word \( w_n \) from \( p(w_n | z_n, \beta) \), a multinomial probability conditioned on the topic \( z_n \).

One needs to specify the number of topics for the model. In the following analysis, we use results from both 10 topics and 20 topics. These are reasonable numbers of policy areas generally discussed in a report. We use the Mallet program, developed by McCallum (2002), to estimate LDA. Estimation proceeds through Gibbs sampling. We monitor Markov Chain convergence in all cases. The chains generally converge before 1,000 iterations. LDA performed very well in generating meaningful clustering of words, and some of these clusters would not be easily detected by a human coder. The vast majority of the topics consist of highly coherent words, and are easily interpretable. The complete term distribution of each topic is available upon request. The appendix following the main text provides the twenty highest-probability words for each topic of the 10-topic LDA, where we have named each topic. Top words for the 20-topic LDA are also available upon request.
We use estimated topic proportions for each report $\hat{\theta}_d$ as the key independent variable for our main analysis. In the context of Gibbs sampling, these are simply the proportions of words assigned to each topic in a given report in the stationary distribution of the Markov Chains.

### B.3 Post-Estimation Diagnostics

One question naturally arises: how noisy is our estimate of topic proportions? The noise in estimation becomes measurement errors, which in turn could potentially introduce attenuation bias in our following empirical analysis. If our estimates are too noisy, we would not be able to say anything about policy priorities. We examine the posterior distribution of the topic proportions to gauge the amount of noise in them.

The theory of MCMC implies that the posterior distribution is the stationary distribution of the Markov Chain. Of the 1,000 sampling iterations, we discard the first 300 iterations as burn-in period, and use a thinning interval of 50. That is, starting from the 300th iteration, we take one sample every 50 iterations. In each sample, we calculate the proportion of words assigned to each topic in the whole corpus. This yields fifteen data points for topic proportions in the whole corpus. Figure A.1 plots these points. The proximity of these points for each given topic is striking, to the extent that they often look like one point. In any case, the range in proportion is smaller than 0.005. This implies that the posterior distribution is very tight, and our estimated topic proportions are very precise. To the extent that the topics are meaningful encapsulations of policy, measurement errors in policy priorities are likely to be very small.
We perform several post-estimation diagnostics to evaluate the qualities of the topics, focusing on two metrics. The first is *coherence*, which measures the tendency for top words in a topic to appear together, and is defined as

$$
\sum_i \sum_{j<i} \log \frac{N(w_j, w_i) + \beta}{N(w_i)}
$$

where $w_i$ is the $i$th ranked word in a given topic, $N(w_i)$ is the number of documents that contain $w_i$, $N(w_j, w_i)$ is the number of documents that contain both $w_j$ and $w_i$, and $\beta$ is a smoothing parameter. A higher coherence indicates that words in this topic are more likely to co-occur. The second is *specificity*, which measures the distinctiveness of the words in a given topic in comparison to a uniform distribution of words, as measured by the Kullback-Leibler divergence between the word distribution of a given topic and the uniform distribution. The more specific a topic is, the greater weight it puts on a somewhat unique set of words (rather than just being a general “background"
Figure A.2 displays the relative positions of the 10 estimated topics in terms of both coherence (x axis) and specificity (y axis). It turns out that the welfare topic is ranked among the highest in both dimensions. This suggests that the welfare topic is one of the best quality topics generated by the 10-topic LDA model.

Figure A.2: Topic Quality: Coherence vs. Specificity

*Note:* This figure illustrates the quality of the welfare topic in terms of both specificity and coherence, in comparison with other nine topics.

### B.4 Estimated Government Work Report Topics

**Topic 1: business attraction and industrial development**

attract businesses, high technology, further, development zones, hundred million US dollars, value added, industrial parks, utilize foreign capital, open up, infrastructure, small and medium enterprises, emerging industries, competitiveness, develop vigorously, manufacturing, commercial
zones, industrialization, major projects, focus projects, service industry

招商引资, 高新技 术, 进 一步, 开发区, 亿 美元, 增值, 工业园 区, 利用外资, 对外 开放, 基础设 施, 中小 企业, 新兴产 业, 竞争 力, 大力 发展, 制造业, 产 业园, 工业 化, 重大 项目, 重 点项目, 服务业

Topic 2: infrastructure and urban development

infrastructure, highway, further, overall planning, treatment plant, ten thousand square meters, project construction, environmental protection, urbanization, preliminary work, ecological environment, small cities, square kilometers, major projects, integration, comprehensive management, coverage rate, construction project, ecological improvement, polluted water treatment

基础设施, 高速公路, 进一步, 总体规划, 处理厂, 万平方米, 工程建设, 环境保护, 城镇化, 前期工作, 生态环境, 小城镇, 平方公里, 重点项目, 一体化, 综合治理, 覆盖率, 建设项目, 生态建设, 污水处理

Topic 3: health, education, and culture

further, compulsory education, family planning, spiritual advancement, public health, health care, carry out thoroughly, primary and secondary schools, broadcasting system, service system, develop vigorously, culture industry, the disabled, carry out extensively, institutional reform, campaign of creating, socialism, mass-line, community health, women and children

进一步, 义务教育, 计划生育, 精神文明, 公共卫生, 医疗卫生, 深入开展, 中小学, 广播电 视, 服务体系, 大力发展, 文化产业, 残疾人, 广泛开展, 体制改革, 创建活动, 社会主义, 群众 性, 社区卫生, 妇女儿童

Topic 4: general economic indicators

gross product, fixed assets, net income, disposable, consumption goods, fiscal revenue, total retail sales, urban residents, value added, economic and societal, unemployment rate, people’s government, growth rate, total income, provincial government, democratic parties, percentage points, household consumption, all sectors of society, rank and file of the armed police

生产总值, 固定资产, 纯收入, 可支配, 消费品, 财政收入, 零售总额, 城镇居民, 增加值, 经济社会, 失业率, 人民政府, 增长率, 总收入, 省政府, 民主党派, 百分点, 居民消费, 各界人士,
Topic 5: real estate and high valued-added service industry

service industry, tourism, develop vigorously, further, tertiary industry, real estate, culture industry, ten thousand people, logistics industry, total income, infrastructure, value added, real estate business, e-commerce, logistics parks, home and abroad, financial institutions, wholesale markets, consumption goods, informatization

服务业, 旅游业, 大力发展, 进一步, 第三产业, 房地产, 文化产业, 万人次, 物流业, 总收入, 基础设施, 增加值, 房地产业, 电子商务, 物流园区, 国内外, 金融机构, 批发市场, 消费品, 信息化

Topic 6: political litany

economic and societal, views on development, further, Three Represents, implementation and solidification, Deng Xiaoping Theory, Eleventh Five-Year, socialism, moderately prosperous society, modernization, provincial government, Twelfth Five-Year, liberalize thoughts, reform and open up, industrialization, urbanization, structural adjustment, deepen the reforms, great standard-bearer, CPC Central Committee

经济社会, 发展观, 进一步, 三个代表, 贯彻落实, 邓小平理论, 十一五, 社会主义, 小康社会, 现代化, 省政府, 十二五, 解放思想, 改革开放, 工业化, 城镇化, 结构调整, 深化改革, 伟大旗帜, 党中央

Topic 7: political institutions

further, administer by law, accountability system, representatives of People’s Congress, Standing Committee, democratic supervision, civil servants, carry out thoroughly, enhance clean politics, democratic parties, CPPCC members, Federation of Industry and Commerce, functions of the government, public service, people’s associations, independents, party ethos and clean politics, democracy and law, public interest, establish and improve

进一步, 依法行政, 责任制, 人大代表, 常委会, 民主监督, 公务员, 深入开展, 廉政建设, 民主党派, 政协委员, 工商联, 政府职能, 公共服务, 人民团体, 无党派人士, 党风廉政, 民主法制, 群众利益, 建立健全
**Topic 8: agriculture modernization**

agricultural produce, commercialization, leading agricultural businesses, develop vigorously, modern agriculture, standardization, further, infrastructure, herding, labor force, organic, structural adjustment, poverty alleviation and development, agriculture and herding, demonstration zone, cooperatives, production capacity, farmers and herders, farmer income, large-scale

农产品, 产业化, 龙头企业, 大力发展, 现代农业, 标准化, 进一步, 基础设施, 畜牧业, 劳动力, 无公害, 结构调整, 扶贫开发, 农牧业, 示范区, 合作社, 生产能力, 农牧民, 农民收入, 规模化

**Topic 9: market reform**

further, institutional reforms, public safety, state-owned enterprises, non-public ownership, state-owned assets, public sector organizations, small and medium enterprises, crack down ruthlessly, comprehensive management, carry out thoroughly, accountability system, investment and financing, financial institutions, private firms, management system, establish and improve, letters and visits affairs, various reforms, misdemeanor and felony

进一步, 体制改革, 社会治安, 国有企业, 非公有制, 国有资产, 事业单位, 中小企业, 严厉打击, 综合治理, 深入开展, 责任制, 投融资, 金融机构, 民营企业, 管理体制, 建立健全, 信访工作, 各项改革, 违法犯罪

**Topic 10: social welfare**

social safety net, endowment insurance, medical insurance, urban and rural residents, urban residents, further, labor force, ten thousand square meters, social security, coverage, migrant workers, new rural, affordable, the disabled, pension, people with financial difficulties, low income, graduates, unemployment rate, urban employees

社会保障, 养老保险, 医疗保险, 城乡居民, 城镇居民, 进一步, 劳动力, 万平方米, 社会保障, 覆盖面, 农民工, 新型农村, 保障性, 残疾人, 养老金, 困难群众, 低收入, 毕业生, 失业率, 城镇职工
C Distribution of Logged Number of Petitions by Year

Figure A.3: Distribution of (Logged) Number of Petitions

Note: This figure illustrates the box and whiskers plot for the (logged) number city-specific petitions in each year. The range of the box represents the inter-quartile range and the horizontal line in the middle represents the median.

D Data on Collective Protests

The Collective Incidents Dataset is a dataset compiled by the institute of sociology at the Chinese Academy of Social Sciences (CASS), one of the major state-sponsored think tanks in China. The project was motivated by the government’s concern about the rising social instability since 2000. Between 2009 and 2013, the project collected on a daily basis information about collective protests and other major mass incidents, and constructed a GIS based dataset, which allows researchers to analyze the spatial distribution and diffusion of mass incidents. Information about protests came from a variety of sources, including internal government documents, state-owned and commercial newspapers, TV reports, websites and social media. Extensive search was conducted on both domestic and foreign sources to minimize potential sampling bias. The dataset contains a rich set

---

1Retrospective collection was also conducted on protests that took place prior to the project start date.
of features regarding protests recorded, such as the course of the event, location, actors, and interaction between actors, causes and consequences, process, number of participants, duration, and the information sources. While this dataset obviously does not exhaust all protests that happened in China during this period, it does provide a decent coverage on the large-scale protests, which are most likely to confound our results by inducing substantive policy changes. For the period of interest (2008-2013), the dataset contains 841 protests, with over 85% of them involving more than a thousand participants. The project was terminated in December 2013.

The other important data source is the *China Strikes* website, which records instances of labor unrest in China based on a variety of sources, including reports from website visitors. This website is dedicated to covering only contentious, collective actions by *workers over workplace issues*. Although this dataset is limited in the scope of protests covered, one of its key advantage is that it is maintained by researchers outside mainland China and therefore less susceptible to censorship issues. For each entry, information about the date, time, location, issues, and identity of participants are provided. Sometimes, the entry will also contain photos and video clips that verify the occurrence of a protest. Between 2008 and 2012, this dataset contains information for a total of 721 incidences.

E Using Semi-automated Content Analysis to Identify Backgrounds of Petitioners

A key challenge in estimating the occupational background of petitioners is that many petitions do not contain direct information about the petitioner’s occupation or residency. To overcome this challenge, we first manually code a large sample of messages on which we can relatively easily determine the petitioners’ class, and then use the semi-automated method proposed by Hopkins and King (2010) (hereafter HK) to estimate the occupational distribution within the entire body of petitioners.

---

2 Currently, this website is maintained by Manfred Elfstrom, a postdoctoral fellow at the Kennedy School of Harvard University.

3 The latest data available on the website are at December 2012.
 petitions. The key assumption here is that petitioners from the similar background tend to have similar grievances and issues that will be captured by similarities in the petition content.

The process of manual coding is as follows: First, we read 2,000 randomly selected sample petitions and then developed the coding book for five largely distinguishable and mutually exclusive classes and groups: rural residents/farmers (农民), workers (工人), cadres (干部), businessman/vendor/managers (企业家/个体户/管理人员), and other professions (其他职业). Below is a sample of keywords for each profession from our coding dictionary

- **Rural residents/farmers (农民):** 农民, 村民, 村里人, 村小组, 农村人, 乡下人, 村居民, 种地, 老农民, 农民工, 打工, 务工人员
- **Workers (工人):** 工人, 技工, 钳工, 工厂里, 厂子里, 工人, 安装工, 车间工, 初级工, 中级工, 高级工, 领班, 维修工, 司机, 师傅, 出租车师傅, 卡车司机, 开车的, 环卫工
- **Cadres (干部):** 干部, 书记, 党委, 主任, 领导, 厅长, 处长, 科长, 局长, 调研员, 巡视员, 公务员, 处级, 科级, 官, 调干, 军队, 部队, 军官, 公安, 警察, 村长, （乡）镇长, 编制内
- **Businessmen/vendors/managers (企业家/个体户/管理人员):** 企业家, 董事长, 总经理, 经理, 总管, 总监, 个体户, 摊贩, 店长, 店主, 厂长, 个体工商户, 总裁, 合伙人, 投资商, 开发商, 老板, 老总
- **Other occupations (其他职业):** 服务员, 快递, 文员, 餐厅, 前台, 教师, 医生, 老师, 记者, 护士, 办事员, 退休, 家庭妇女, 下岗的, 保安

Next, 12 trained research assistants read each petition and categorized it into different social groups. A total of 74,867 petitions were identified with explicit social identities. Within the identified petitions, 29.5% were submitted by rural residents, 13.5% of the petitions were created by workers. Cadres, businessmen, and other professions contributed to 6.5%, 4.0%, and 18.5% of the petitions, respectively.

Because the sample of messages with identifiable classes is unlikely to be a random sample, we cannot generalize the occupational distribution in the identified sample to the full sample. How-
ever, the HK method does not require the training set to be a random sample, as long as language patterns in the training set is the same as in the population. In this situation where human coders have difficulty identifying messages, classification algorithms are unlikely to classify individual messages with much accuracy. One practical challenge of implementing the HK method, however, is that ReadMe, the R package that performs the estimation procedure, has a memory constraint and therefore cannot handle datasets that are too big. To circumvent this problem, we break the data by quarter, conduct separate estimations on each of the smaller dataset, and report the average in the main text. Figure A.4 displays in greater detail the quarterly results. We find the distribution to be quite stable over time.

Figure A.4: Estimated Occupational Proportion By Quarter

We conduct ten-fold cross validation on the set of 74,867 identified messages. Figure A.5 shows the results. Each point shows true (hand-coded) and estimated (HK-generated) proportion of one occupation in one cross validation trial. We can see from the figure that the points fall quite closely around the 45 degree line. Estimation quality is slightly lower for intermediate proportions. Overall this shows that the HK model does a very good job in estimating occupation proportions.
F  Robustness Checks

F.1  Alternative Estimation Strategy

We conduct three sets of robustness checks of the baseline results. First, we note that the dependent variable is a proportion that takes value between 0 and 1. One potential concern is thus that our results might be driven by the bounded nature of the dependent variable. Visual inspection of the distribution of the DV, however, suggests that it follows a quite “normal” shape. To further address this possibility, we follow the advice by Papke and Wooldridge (1996) to adopt two additional estimation strategies: fractional logit and beta regression, both of which are designed to deal with proportional dependent variables. The estimation results are presented in the first two columns of Table A.2, and the main coefficient of interest remains positive and statistically significant.

Another concern that our dependent variable tends to be serially correlated because of each
year’s GWR will devote some sections to discuss previous year’s accomplishments. Failing to take into account the autocorrelated structure may result in smaller standard errors and false rejection of the null (no effect). Our main strategy to address this problem is to cluster the standard error at the city level for all regression results reported in the main text. To further address this issue, we also re-estimate the baseline regression using incremental changes in welfare topic (\textit{Welfare topic share at }t + 1 - \textit{Welfare topic share at }t) as the dependent variable. As shown in Column 3 of Table A.2, the main result remains robust when using this modified dependent variable.

<table>
<thead>
<tr>
<th>Table A.2: Robustness Checks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welfare Topic at (t + 1) (10-topic)</td>
</tr>
<tr>
<td>(1) Beta regression</td>
</tr>
<tr>
<td>Log petitions</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>City and year fixed-effects</td>
</tr>
<tr>
<td>Province-specific trends</td>
</tr>
<tr>
<td>Adjusted (R^2)</td>
</tr>
<tr>
<td>Number of Cities</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

\textbf{Note}: This table shows results from several robustness checks. The first two columns present results from using alternative estimation strategies that take into account the proportional nature of the dependent variable. The last column uses incremental changes in the original dependent variable as a way to address potential serial correlation problems. The results are substantive the same as the baseline. Robust standard errors clustered at city level are reported in parentheses.

\( ^* p < 0.1, ^{**} p < 0.05, ^{***} p < 0.01 \) (two-tailed test)

\section*{F.2 Dealing with Reverse Causality}

We take extra effort to address the issue of reverse causality, which can be a challenge to our causal claim. It is possible, for example, that when local governments place greater emphasis on social welfare issues, lower-class citizens are mobilized/encouraged to participate through various channels because they now perceive a greater chance of receiving useful responses. To test this possibility, we run two additional regressions with the timing of the dependent variable set at \(t\)
and \(t - 1\), respectively.\(^4\) If the mobilization story is true, then we would expect content of earlier GWRs to have a much stronger association with petition intensity than later ones. However, results presented in Table A.3 suggest that this is not the case: neither the current or the previous year’s GWR welfare topic appears to be significantly correlated with the current year’s volume of online petition.

Table A.3: Alternate the Timing of the Dependent Variable

<table>
<thead>
<tr>
<th>Welfare Topic (10-topic) at ...</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t)</td>
<td>-0.0007</td>
<td>-0.0000</td>
</tr>
<tr>
<td>(t - 1)</td>
<td>(0.0018)</td>
<td>(0.0019)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>City and year fixed-effects</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Province-specific trends</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Adjusted (R^2)</td>
<td>0.23</td>
<td>0.14</td>
</tr>
<tr>
<td>Number of Cities</td>
<td>299</td>
<td>297</td>
</tr>
<tr>
<td>Observations</td>
<td>1656</td>
<td>1614</td>
</tr>
</tbody>
</table>

Note: This table shows results from altering the timing of the dependent variable. The results suggest that earlier welfare topic proportions in GWRs is not significantly associated with subsequent petition intensity. This finding helps mitigate some of the concerns about reverse causality. Robust standard errors clustered at city level are reported in parentheses.

\(^* p < 0.1\), \(^* * p < 0.05\), \(^* * * p < 0.01\) (two-tailed test)

As another way to check whether the effect of online participation is endogenous to government policies, we estimate several regressions that include both the current and past welfare topic shares for the provincial government as well as the past welfare topic share for the city governments. Due to the well-known Nickell bias (Nickell 1981), we do not include city fixed-effects in these regressions to avoid incidental parameter problem. The results are presented in Table A.4. The first two columns include lagged welfare topic share for city GWR \((t \text{ and } t - 1)\), and Columns 3 through 5 include the welfare topic shares in provincial GWRs at \(t + 1\), \(t\), and \(t - 1\), respectively. Finally, the last column presents a model that includes all the preceding controls. Through all the models,

\(^4\)More specifically, GWRs at year \(t\) correspond to the reports that are published at the beginning of a year (usually January or February).
we see that the coefficient estimate for our main variable of interest, Log Petition remains, remains highly robust and the magnitude remains virtually unchanged with the inclusion of these additional controls. These results lend us further confidence that the observed effect of online participation is unlikely to be driven solely by changes in local government policies.

Table A.4: Controlling for Welfare Topic Share at the Higher Level or from Earlier Periods

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log petitions</td>
<td>0.0032***</td>
<td>0.0034***</td>
<td>0.0034**</td>
<td>0.0034**</td>
<td>0.0034**</td>
</tr>
<tr>
<td>Welfare topic share at t (city GWR)</td>
<td>0.3167***</td>
<td>(0.0111)</td>
<td>(0.0012)</td>
<td>(0.0014)</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>Welfare topic share at t – 1 (city GWR)</td>
<td>0.2042***</td>
<td>(0.0320)</td>
<td>(0.0313)</td>
<td>0.2647***</td>
<td>(0.0287)</td>
</tr>
<tr>
<td>Welfare topic share at t + 1 (provincial GWR)</td>
<td>0.0434</td>
<td>(0.0332)</td>
<td>(0.0332)</td>
<td>-0.066</td>
<td>(0.0433)</td>
</tr>
<tr>
<td>Welfare topic share at t – 1 (provincial GWR)</td>
<td>-0.0393</td>
<td>(0.0338)</td>
<td>(0.0338)</td>
<td>-0.0409</td>
<td>(0.0429)</td>
</tr>
<tr>
<td>Welfare topic share at t + 1 (provincial GWR)</td>
<td>-0.0066</td>
<td>(0.0433)</td>
<td>(0.0433)</td>
<td>-0.0632</td>
<td>(0.0429)</td>
</tr>
</tbody>
</table>

Year fixed-effects ✓ ✓ ✓ ✓ ✓ ✓
Province-specific trends ✓ ✓ ✓ ✓ ✓ ✓
Adjusted R² 0.32 0.27 0.24 0.26 0.24 0.32
Number of Cities 297 297 300 300 300 295
Observations 1624 1580 1685 1685 1685 1553

Note: This table shows regression results from including a variety of controls for the welfare topic share in provincial-level GWR or that in past city-level GWRs. The results indicate that our main coefficient of interest is highly robust to the inclusion of these potential confounders, suggesting that the effect of online participation is not endogenous to governments’ prior policy orientations.
Robust standard errors clustered at city level are reported in parentheses.
* p < 0.1, ** p < 0.05, *** p < 0.01 (two-tailed test)

F.3 Effect of Participation over Time

We also investigate the possibility that our results were driven by rared events that happened a few special years. To do so, we estimate a model where the participation variable is interacted with a set of year dummies. It is clear that the effect is in fact quite stable over time. All estimates are statistically significant at 90% level or higher with the exception of 2011. This pattern lends us further confidence that the result is not driven by the choice of sample period.
F.4 Permutation Inference

Finally, we reassess the statistical significance of our main estimate through a permutation test. The basic idea of the test is to simulate a null distribution of the coefficient interest (effect of petition on welfare topic proportion) by running a series of regressions where the outcome (welfare topic proportion) is randomly assigned to treatment conditions (petition level), and then use the simulated distribution as the benchmark to assess the likelihood of observing an effect at least as extreme as the actual estimate. In performing the permutation, we impose two additional constraints to preserve some basic feature of the dataset. First, we take into account the nested structure by limiting reshuffles to cities within a province. Second, we also try to preserve the temporal relationship in our variables by reshuffling not individual observations but the entire sequence of observations between 2008 and 2013. In other words, for each city-year spell, the level of petition that will be assigned to it will come from a city within the same province in the same year. We perform 10,000 permutation and plot the distribution in Figure A.7. The red vertical line indicates the position of the value of the actual estimate. It turns out that only about 3.4% of the permutations yield an
estimate that is greater than the actual result. The permutation result thus reconfirms the statistical significance of our coefficient of interest.

Figure A.7: Permutation Inference

![Permutation Inference](image)

\[ p \text{ value} = 0.034 \]

\[ 0 \]

\[ 100 \]

\[ 200 \]

\[ 300 \]

\[ -0.004 \]

\[ 0.000 \]

\[ 0.004 \]

\[ G \text{ Mapping Responsiveness by Province} \]

We also estimate province-specific policy responsiveness using the following equation:

\[
\text{Welfare Topic Share}_{it+1} = \alpha \log \text{Petitions}_{it} + \delta_p \log \text{Petitions}_{it} \times \text{Province}_p + X_{it} \beta + \phi_i + \tau_t + \epsilon_{it}.
\]

The level of policy responsiveness for province \( p \) is thus \( \alpha + \delta_p \). We plot this quantity in the following map.
Figure A.8: The Geography of Policy Responsiveness

Point estimates for policy responsiveness

- No data
- $[-0.01,0)$ (<20%)
- $[0,0.002)$ (20−40%)
- $[0.002,0.005)$ (40−60%)
- $[0.005,0.006)$ (60−80%)
- $[0.006,0.012]$ (>80%)
References


