Econometrica Supplementary Material

SUPPLEMENT TO "SOCIAL MEDIA AND COLLECTIVE ACTION IN CHINA"

(*Econometrica*, Vol. 92, No. 6, November 2024, 1993–2026)

BEI QIN
Department of Accountancy, Economics and Finance, Hong Kong Baptist University

DAVID STRÖMBERG
Economics Department, University of Stockholm

YANHUI WU
HKU Business School, University of Hong Kong

APPENDIX: SUPPLEMENTAL APPENDIX

A.1. *Specification*

To examine the problem of correlated and contextual effects, we start with the formulation similar to Moffitt (2001). Suppose that we have $g = 1, \dots G$ groups and that there are only two cities ($i = 1, 2$) per group. Let y_{ig} be the outcome variable of interest for city i in group g, x_{ig} be an observed characteristic of i, which in our case is the lagged outcome, and v_{ig} be an unobservable characteristic. Assume that the model that we wish to estimate is

$$y_{1g} = \alpha_0 + \alpha x_{1g} + \beta x_{2g} + v_{1g},$$
$$y_{2g} = \alpha_0 + \alpha x_{2g} + \beta x_{1g} + v_{2g}.$$

In particular, we are interested in β, which captures how lagged events spread from city $1g$ to connected city $2g$, and vice versa.

A concern is that v_{1g} and v_{2g} are correlated, for example, because of correlated and contextual effects. The correlated effects are typically parametrized by an unobserved group-specific unobserved shock, δ_g, which is part of the error and for which

$$E(\delta_g x_{ig}) \neq 0.$$

The contextual effects are typically parameterized by including a term γx_{2g} in the error term v_{1g}. In our case, contextual effects could arise from events spreading through communication channels that are correlated with the social media network, such as cell-phone use. Thus, the error term, v_{1g}, can be decomposed as

$$v_{1g} = \gamma x_{2g} + \delta_g + \varepsilon_{1g}.$$

By the same token, v_{2g} is similarly defined.

Bei Qin: beiqin@hkbu.edu.hk
David Strömberg: david.stromberg@ne.su.se
Yanhui Wu: yanhuiwu@hku.hk

We are grateful to Davide Cantoni, Chun-Fang Chiang, Peter Hull, Ruixue Jia, Adam Szeidl, Jaya Wen, David Yang, Noam Yuchtman, and numerous seminar and conference participants for helpful comments. This study was approved by the Regional Ethics Review Board of Stockholm, 2017/1288-31. The project has received funding from the European Research Council (ERC project 742983) and the Swedish Research Council (diary 2016-03220).

In this model, $\gamma \neq 0$ implies contextual effects and $\delta_g \neq 0$ implies correlated effects. Clearly, β is not identified in the presence of these effects.

A.1.1. *Setup*

We estimate our model using a panel of N Chinese cities at daily frequency, t. Let y_{it} be a dummy variable indicating the occurrence of a protest. Suppose that the probability of a protest in city i on day t, $\Pr(y_{it})$ depends on the number of people who are informed about protests y_{jt-1} in another city j at time $t-1$. Let f_{ijt} be the number of people in city i who read posts from users in city j at time t and assume that the number of people who learn about the protest is $f_{ijt}y_{jt-1}$. We wish to model how protests spread from city j to city i as a function of $f_{ijt}y_{jt-1}$.

We illustrate the identification assumptions with the following simple example. There are two periods (pre-Weibo and post-Weibo) and two groups of cities (N, C) that may be affected by a protest in another city j on a previous date. Cities in group N are never connected via social media to city j, whereas cities in group C are connected via social media to city j in the post-Weibo period only. Social media connections increase the probability that a protest spreads from j to other cities. In both periods, there are some days with protests in location j, and some days without such protests. Let y_{ji-1} be an indicator variable for a protest in city j at $t-1$, f_i be an indicator variable for city i belonging to group C and P_t be an indicator variable for the post-Weibo period. Then we have

$$y_{it} = \delta + \alpha y_{it-1} + \beta^p f_i y_{jt-1} + v_{it}, \tag{A1}$$

where β^p measures the effect of events spreading through the social media network with $\beta^0 = 0$ in the pre-Weibo period $(p = 0)$ and $\beta^1 = \beta$ in the post-Weibo period $(p = 1)$. Further, let

$$v_{it} = \gamma f_i y_{jt-1} + f_i \delta_g + \varepsilon_{it}, \tag{A2}$$

where $\gamma f_i y_{jt-1}$ captures the contextual effects and $f_i \delta_g$ captures the correlated effects.

A.1.2. *Triple-Differences Estimator*

Consider the following triple-differences equation in the linear probability model, with outcome variable y_{it} being a protest in city i at time t,

$$\begin{aligned} y_{it} = {} & \beta_0 + \beta_1 f_i + \beta_2 y_{jt-1} + \beta_3 P_t \\ & + \beta_4 f_i y_{jt-1} + \beta_5 f_i P_t + \beta_6 P_t y_{jt-1} \\ & + \beta P_t f_i y_{jt-1} + \varepsilon_{it}. \end{aligned} \tag{A3}$$

To simplify the exposition, we drop from the equation the term αy_{it-1}, which captures within-city spread. The conditional mean function $E[y_{it}|f_i, y_{jt-1}, P_t]$ can take on eight values, and the model is saturated because it has eight parameters. The first row of Equation (A3) contains the three main effects and the constant, the second row contains the three two-way interactions, and the third row contains the triple interaction. This model allows for the probability of a protest in city i to be larger in connected cities (β_1) than in nonconnected cities, for example, because of correlated effects (absorbing $f_i\delta_g$) in the post-Weibo period (β_3), and additionally higher for connected cities in the post-Weibo period (β_5). Protests are also allowed to spread across all cities (β_2), and differentially so in the post-Weibo period (β_6). In particular, note that protests in city i may be more likely to occur immediately after a protest in a connected city j in the pre-Weibo period (β_4), such as because of contextual effects (absorbing $\gamma f_i y_{jt-1}$). The coefficient β captures the difference in the spread of protests between connected and nonconnected cities before and after Weibo entry.

Under the standard OLS assumption, $E[\varepsilon_{it}|f_i, y_{jt-1}, P_t] = 0$. For compactness, we use the notation

$$E[Y|f_i = 1, y_{jt-1} = 1, P_t = 1] = Y_{C,S,\text{Post}},$$

where C denotes a connected city ($f_i = 1$), S denotes lagged protests in city j ($y_{jt-1} = 1$), and Post denotes $P_t = 1$. Similarly, let N denotes nonconnected city ($f_i = 0$), W denote a day without a protest in city j, $y_{jt-1} = 0$, and Pre denote that $P_t = 0$. It is straightforward to show that $\beta = \delta_C - \delta_N$, where

$$\delta_C = (Y_{C,S,\text{Post}} - Y_{C,W,\text{Post}}) - (Y_{C,S,\text{Pre}} - Y_{C,W,\text{Pre}})$$

is the difference in city C between the pre-Weibo and post-Weibo periods in the difference in protest incidence when there is a protest or not in city j, and

$$\delta_N = (Y_{N,S,\text{Post}} - Y_{N,W,\text{Post}}) - (Y_{N,S,\text{Pre}} - Y_{N,W,\text{Pre}})$$

is the equivalent difference in city N.

By rearranging terms, we can also write the estimator as $\beta = \delta_1 - \delta_0$, where

$$\delta_1 = (Y_{C,S,\text{Post}} - Y_{C,W,\text{Post}}) - (Y_{N,S,\text{Post}} - Y_{N,W,\text{Post}})$$

is the difference between connected and nonconnected cities in the difference between days with and without protests in the post-Weibo period, and

$$\delta_0 = (Y_{C,S,\text{Pre}} - Y_{C,W,\text{Pre}}) - (Y_{N,S,\text{Pre}} - Y_{N,W,\text{Pre}})$$

is the equivalent difference in differences in the pre-Weibo period.

A.1.3. *Parallel-Trends Assumption*

Below, we articulate the identification conditions for the triple-difference estimator we have derived under the potential outcomes framework.

Let $E[Y_1|C, S, \text{Post}]$ denote the expected outcome in state (C, S, Post) in the case that a city is connected and $E[Y_0|C, S, \text{Post}]$ denote the expected outcome in the counterfactual case in which the city is not connected. The estimate β_{TT} captures the causal effect of being connected to a city with a protest in the post-Weibo period for the connected cities, formulated as follows:

$$\beta_{TT} = E[Y_1|C, S, \text{Post}] - E[Y_0|C, S, \text{Post}].$$

Our estimator is

$$\begin{aligned}
\beta &= \delta_C - \delta_N \\
&= \big(E[Y_1|C, S, \text{Post}] - E[Y_0|C, W, \text{Post}]\big) - \big(E[Y_0|C, S, \text{Pre}] - E[Y_0|C, W, \text{Pre}]\big) \\
&\quad -\big(E[Y_0|N, S, \text{Post}] - E[Y_0|N, W, \text{Post}]\big) - \big(E[Y_0|N, S, \text{Pre}] - E[Y_0|N, W, \text{Pre}]\big).
\end{aligned}$$

The condition for $\beta = \beta_{TT}$ is that absence treatment, $\delta_C = \delta_N$. Since the only term that contains treatment in δ_C is the first term, $E[Y_1|C, S, \text{Post}]$, the parallel-trend condition is

$$\begin{aligned}
&\big(E[Y_0|C, S, \text{Post}] - E[Y_0|C, W, \text{Post}]\big) - \big(E[Y_0|C, S, \text{Pre}] - E[Y_0|C, W, \text{Pre}]\big) \\
&= \big(E[Y_0|N, S, \text{Post}] - E[Y_0|N, W, \text{Post}]\big) \\
&\quad - \big(E[Y_0|N, S, \text{Pre}] - E[Y_0|N, W, \text{Pre}]\big). \qquad \text{(A4)}
\end{aligned}$$

A.1.4. *Contextual and Correlated Effects*

Our estimator differences out the contextual and correlated effects in Equations (A1) and (A2). The correlated effects, $f_i\delta_g$, are differenced out because both δ_C and δ_N are constructed solely from the within-group differences.

The contextual effects are partialed out because they are contained in both δ_1 and δ_0, and $\delta_1 - \delta_0 = (\beta + \gamma) - \gamma = \beta$. The parallel-trends assumption in Equation (A4) implies that the difference-in-differences captured by δ_0 would have been the same in the post-Weibo period, in the counterfactual case that the social media network did not exist,

$$\big(E[Y_0|C, S, \text{Post}] - E[Y_0|C, W, \text{Post}]\big) - \big(E[Y_0|N, S, \text{Post}] - E[Y_0|N, W, \text{Post}]\big) = \gamma.$$

We cannot test the assumptions directly because we do not observe $E[Y_0|C, S, \text{Post}]$. However, we can estimate γ separately for each half-year, b, using

$$\widehat{\gamma}_b = (\overline{Y}_{C,S,\text{Pre}} - \overline{Y}_{C,N,\text{Pre}})_b - (\overline{Y}_{N,S,\text{Pre}} - \overline{Y}_{N,N,\text{Pre}})_b$$

to see whether this exhibits a trend in the pre-Weibo period.

A.1.5. *Multiple Cities*

In the real-world data that we use, protests occur in all cities, and social media connections in the post-Weibo period are of different strengths. Let f_{ij} be the strength of city i's connection to city j. Now, we generalize the above model by linearly adding the contributions described in Equation (A3) from each city j,

$$\begin{aligned}
y_{it} = {} & \beta_0 + \beta_1\overline{f}_i + \beta_2(y_{t-1} - y_{it-1}) + \beta_3 P_t \\
& + \beta_4 s_{it-1} + \beta_5\overline{f}_i P_t + \beta_6 P_t(y_{t-1} - y_{it-1}) \\
& + \beta P_t s_{it-1} + \varepsilon_{it},
\end{aligned}$$

where $\overline{f}_i = \sum_j \overline{f}_{ij}$, $y_{t-1} = \sum_j y_{jt-1}$, and $s_{it-1} = \sum_{j\neq i} f_{ij} y_{jt-1}$. To allow for more flexibility, we will estimate an equation of the form

$$y_{it} = \alpha^p y_{it-1} + \beta^p s_{it-1} + \delta_i^p + \delta_t + \varepsilon_{it}, \qquad \text{(A5)}$$

where superscripts indicate the pre-Weibo and post-Weibo periods. The city-by-period fixed effects absorb the following terms:

$$\begin{aligned}
\delta_i^0 &= \beta_0 + \beta_1\overline{f}_i \\
\delta_i^1 &= \beta_0 + \beta_1\overline{f}_i + \beta_3 P_t + \beta_5\overline{f}_i P_t.
\end{aligned}$$

The date-fixed effects and the lagged outcome in city i, y_{it-1}, interacted with the period-fixed effects p_t, absorb the following terms:

$$\beta_2 y_{t-1} + \beta_3 P_t + \beta_6 P_t y_{t-1} + (-\beta_2 - \beta_6 P_{t2}) y_{it-1} = \delta_t + \alpha^0 y_{it-1} + \alpha^1 y_{it-1},$$

where $\delta_t = \beta_2 y_{t-1} + \beta_3 P_t + \beta_6 P_t y_{t-1}$, $\alpha^0 = -\beta_2$, and $\alpha^1 = -\beta_2 - \beta_6 P_{t2}$. To simplify the exposition, we omit the term involving the within-city spread, αy_{it-1}, from Equation (A1) when we constructed Equation (A3). To include this spread, α should be added to α^0 and α^1. Finally, our specification includes s_{it-1}, interacted with period-fixed effects p_t. This absorbs the terms

$$\beta_4 s_{it-1} + \beta P_t s_{it-1} = \beta^0 s_{it-1} + \beta^1 s_{it-1},$$

where $\beta^0 = \beta_4$ and $\beta^1 = \beta_4 + \beta$. After adding controls, Equation (A5) corresponds to Equation (2) in the main manuscript.

A.2. *Monte Carlo Simulations*

A.2.1. *Stationarity*

In our setting, the stationarity of the process of protests or strikes is not simply an econometric issue. Whether these processes are stable or exploding is likely to be a core concern of an authoritarian regime (the Chinese central government in our context). Consider the following simple first-order serial autoregressive process: $y_t = \rho y_{t-1} + \varepsilon_t$.

This process is stationary if $|\rho| < 1$. If $\rho > 1$, then, on average, each y_t is larger than the past y_{t-1}, and in expectation, y_t grows exponentially over time. We can investigate whether a process is on such an exploding path by simulating outcomes generated from drawing random shocks from the distribution of ε_t, and iteratively computing sequences of y_t. Our setting is more complex than this simple example because of the network structure, multiple time lags, and discrete outcomes, but the principle remains the same.

In a dynamic spatial panel data model, stationarity depends on the parameters of the model as well as on the spatial weight matrix, which determines the amount of autocorrelation, or feedback, in the process. For the location's own autoregressive term and the distance weighted term, this feedback is constant over time because α, γ, and the distance matrix D are constant. However, more intensive use of social media will increase the feedback, and each individual row in the forwarding matrix, F, does not sum up to one. This implies that the marginal effect of a change in y_{t-1} on the probability of an event differs across location and time. In the linear model, the average effect on a particular date t equals $\beta\overline{f}_{t-1}$, where $\overline{f}_{t-1}$ is the average row sum across locations on that date. The maximum of such a row sum is 10.5, an order of magnitude larger than the average row sum across all dates, t, which is normalized to one. This implies that the sufficient conditions for stationarity are not fulfilled for the linear model. In other words, stationarity requires a concave function for high values of s_{it-1}, as we demonstrate below.

In the baseline estimation, we assume that s_{it-1} enters linearly in Equation (1). Now, we investigate this assumption using a nonparametric least squares regression (Cattaneo, Crump, Farrell, and Feng (2024)). Specifically, we estimate the nonparametric conditional mean function $h(\cdot)$ using the following specification:

$$y_{it} = \alpha y_{it-1} + h(s_{it-1}) + \gamma d_{it-1} + \theta_0 w_{it} + \theta' x_{it} + \delta_i + \delta_t + \varepsilon_{it}. \tag{A6}$$

Figure 5 in the paper plots the nonparametric conditional mean function (shown by the dots), estimated using the specification in Equation (1), together with three parametric approximations: a linear approximation, a logarithmic ($\log(5s_{it-1}+1)$) approximation, and a fifth-order polynomial approximation. The conditional mean function is approximately linear for most of the support, and all three approximation functions yield very similar results for the estimated average marginal effects.[1]

However, as shown in Figure 5, when an event wave grows above a certain extent, the marginal spread effect of an additional event (the slope of the curve) falls. Thus, for sufficiently high values of s_{it-1}, there is no spread of events through social media. This causes the magnitude of y_t to fall rapidly in the right tail. Because of this feature, the process is never on an exploding path when we use either the fifth-order polynomial or the log approximations of the conditional mean functions for the data-generating process. Given that the log-model only has one parameter, and is sufficiently concave to avoid exploding paths, we use this functional form in most of the simulations.

A.2.2. *Nickell Bias*

The Nickell bias arises in dynamic panels with fixed effects. In our estimation, we presume that the Nickell bias is small because the T in our panel model is very large. Nevertheless, we run a set of Monte Carlo simulations to assess the Nickell bias in the estimated coefficients of our baseline model. We first estimate the parameters α, β, γ, δ_t, and δ_i from a regression specified as in Equation (1), using the logarithmic function described above and without Weibo penetration and controls (for the sake of simplicity). Next, we generate data using the same model with the estimated parameters, adjusted such that $\delta_t + \delta_i \geq 0$. Then we estimate the model parameters $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma})$ on the simulated data. We repeat this procedure 100 times. Figure A.4 plots the distribution of t-statistics of coefficients α, β, and γ in Equation (1) against the standard normal density. The bias is very small, as evident from the negligible difference between the true and the mean estimated β for both protests and strikes.

In Section 5, we report the estimates of a model including fixed effects for arbitrary time-constant spread across locations. We also run a set of Monte Carlo simulations to assess the Nickel bias in this model. Specifically, we use the baseline model for the data-generating process and then estimate the interaction-fixed-effects model using the simulated data. Figure A.5 depicts the distribution of β-estimates from the Monte Carlo simulations. The graphs to the left show the results from estimations without interaction-fixed effects (Equation (1)), corresponding to specifications in columns (1) and (4) of Table I. The graphs to the right are based on the regressions with interaction-fixed effects, corresponding to the specifications used in columns (3) and (6). The blue line shows the true coefficient used in the data-generating process. The red line represents the mean coefficients from the estimated on the simulated data. These graphs show a bias of 0.011 for both protests and strikes.

We also test for the possibility of autocorrelated errors on the simulated data in which the autocorrelation in errors is absent by construction. The test for autocorrelation in the baseline model verifies this, although it slightly overrejects the nonautocorrelation hypothesis. The model with interaction-fixed effects can control for the same pattern as the baseline model, but many interaction-fixed effects are imprecisely estimated. Removing slightly incorrect autocorrelated terms generates autocorrelated errors. Hence, it is not

[1]This is evident from results using the logarithmic function in the Appendix Tables A5 and A6, in comparison with our baseline results (Tables I and II). The logarithmic model uses the conditional mean function $\ln(5s_{it-1} + 1)$. The estimates of β are statistically significant across all specifications and the implied marginal effects on event probabilities are similar to those estimated using the linear model.

surprising that the autocorrelation test for the model with interaction-fixed effects incorrectly rejects no autocorrelation in our data, which is simulated with no autocorrelation.

Presumably, autocorrelation would disappear as the sample size grows to infinity, because the coefficients are consistently estimated and would converge to the data generation process. The lack of autocorrelation in the errors in Table I shows that there is no significant spread of protests other than what is captured in the baseline model. Although the specifications used in columns (3) and (6) of Table I generate autocorrelated errors in small samples, they will not severely bias the estimated coefficients, as shown in Figure A.5.

A.2.3. *Mechanical Zero Effect*

The estimated social media effect on event spread is not closely related to the number of events in each period. For example, there are 50% more protests in period 2 than period 1 (1516 compared with 1037), but the estimated spread in period 2 is smaller than in period 1. Similarly, while there are six times as many strikes in period 2 than in period 1 (7857 compared with 1365), the estimated spread effect is not significantly different. The lack of, or even a negative, relationship between the number of events and the size of the estimated effects suggests that there is no mechanical relationship between these two variables.

Concerns may remain that the smaller number of events in the pre-Weibo period would mechanically reduce the estimated effects. To further explore this, we simulate event data using a process that matches the observed event frequencies within each period, with the propagation of protest waves across cities being equally strong in the pre-Weibo and post-Weibo periods. Then we estimate the effects on the simulated data and test whether the coefficients in the pre-Weibo period are significantly lower than in the post-Weibo period.

To implement this research design, we first run a set of Monte Carlo simulations to generate event data for the period from 2010 to 2013. To simulate data with lower event frequency for the pre-Weibo period (2006–2010), we use the post-Weibo data-generating process, and then sample sequences starting from a random date and continuing until the same number of events as in the real data are reached. We retain simulated events $\widetilde{y}_{it}$ in the sampled sequences but drop all other events (i.e., y_{it} is set to zero).

Then we reestimate the model of Table II based on these simulated data. Table A4 shows the results. Despite the much smaller number of events in the pre-Weibo period, the average estimated β-coefficient is only slightly lower in the pre-Weibo period than in the post-Weibo period (0.10 compared with 0.12). When we test whether the coefficient in the pre-Weibo period is lower than in the post-Weibo period at the 5% significance level, the hypothesis is rejected in 8% of the simulations.

A.2.4. *Magnitude of Event Waves*

In this section, we estimate the effect of social media on the size of an event wave, measured by the number of essentially simultaneous events across multiple cities. We are not only interested in the mean number of events but also in the likelihood of very large event waves. This requires that we specify some additional details about the dynamic process of event waves.

In addition to the effects through social media, measured by s_{it-1}, the propagation of protests may depend on the size of event waves in the real world, measured by the number of essentially simultaneous events, $y_{t-1} = \sum y_{it-1}$. On the same date, y_{t-1} is constant across all cities, and hence is absorbed by the date fixed effect, δ_t, in Equation (1). The specification including date-fixed effects is preferred when the purpose is to identify the marginal effect of s_{it-1}, because it flexibly controls for all time-constant heterogeneity, including that through y_{t-1}. However, this specification does not explicitly model the dynamic effects through y_{t-1}. Therefore, when the objective is to explore the size of event waves, we need to bring the effect of y_{t-1} to the scene. We do so by adding y_{t-1} to Equation (1) and replacing the date-fixed effects with fixed effects for running months. More specifically, we use the parameters $(\widehat{\alpha}, \widehat{\gamma})$ and the fifth-order polynomial $\widehat{h}_5(s_{it-1})$, estimated as described above, which are obtained from the specification in Equation (1) that includes date fixed effects (because this model most convincingly identifies these parameters). Then we estimate the parameters δ_i, ρ, δ_m using the following specification:

$$y_{it} = \widehat{\alpha} y_{it-1} + \widehat{h}_5(s_{it-1}) + \widehat{\gamma} d_{it-1} + \delta_i + \rho y_{t-1} + \delta_m + \varepsilon_{it}, \tag{A7}$$

where the date fixed effects, δ_t, are replaced with date-constant variables: the total number of events in the past two days, y_{t-1}, and the month-fixed effects, δ_m. The estimate of $\widehat{\rho} \approx -0.002$ implies that the probability of an event falls by 0.02 for every 10 additional events that occurred in the past two days.

Then we use the above model to simulate event data under two scenarios, with and without social media. In both scenarios, we use the estimated probabilities that events erupt independently of other events, $\widehat{\delta}_i + \widehat{\delta}_m$, and we keep the within-city and geographical and aggregate spread, $(\widehat{\alpha}, \widehat{\gamma}, \widehat{\rho})$ at the estimated levels. In one scenario, we allow events to spread through social media using the estimated nonparametric conditional mean function, $\widehat{h}_5(s_{it-1})$. In the other scenario, we do not allow for event spread over social media by omitting the function $\widehat{h}_5(s_{it-1})$ from the data generating process. We simulate the data 1000 times, obtaining 1000 possible histories of protest events in our city panel up to 2013. For each simulation and year, we compute the maximum number of cities with protests within the same week and the corresponding maximum share of population in the cities affected by protests. The results are shown in Figure 6 and discussed in Section 6.3 of the paper.

A.3. *Censorship*

Censorship and Aggregate Retweets. In 2015, we checked which posts in a subsample of posts in our data set remained online. Based on this deletion rate at the regional level, we construct a measure of local censoring intensity. Given the literature on censorship in China, it seems unlikely that censorship significantly affects the number of retweets on nonsensitive topics. We verify this by regressing our time-constant measure of social media connections, f_{ij}, on the average share of deleted posts in city i and city j. All of the variables are standardized, and standard errors are clustered by city i and city j. Table A9 reports the results. It is clear that f_{ij} is not significantly related to the share of deleted posts in either the city of the tweeting user (j) or the city of the retweeting user (i). Column (2) adds city-level controls and Weibo penetration in 2012. The coefficients of interest are nonsignificant, and the implied magnitudes are small. Column (3) interacts the share of deleted posts with Weibo penetration, because the effect on total retweets should be increasing in Weibo use. The coefficient of this interaction term is small and statistically insignificant.

Heterogeneity Analysis. We further investigate whether the spread of protests and strikes is correlated with our measure of local censorship intensity. First, censorship may affect outgoing messages from the city of event occurrence. Suppose that the event y_{jt-1} in city j at time $t-1$ causes a burst of tweets, a share of which, c_j, are censored. To capture this type of effect, we construct the variable $\sum_{i \neq j} f_{ij} c_j y_{jt-1}$ where c_j is the measure of the average share of deleted posts in city j. However, the variables $\sum_{i \neq j} f_{ij} c_j y_{jt-1}$ and $\sum_{i \neq j} f_{ij} y_{jt-1}$ are highly collinear, making it impossible to separately identify the effects from these two variables. Such multicollinearity arises because many events occur in cities with deletion rates close to the mean deletion rates, and thus the variance in the weights c_j is small, making the two variables highly correlated.

Second, we consider the situation in which censorship may prevent social media users in affected cities (to which the events could spread) from reading or retweeting incoming event-related tweets from other cities. We capture this by using the variable $c_i \sum_{i \neq j} f_{ij} y_{jt-1}$ where c_i is the share of deleted posts in city i. As shown in Table A10, neither the spread of protests nor the spread of strikes is significantly different in cities with different levels of censorship.

A.3.1. *Censorship, Content Exposure, and Conclusions About Mechanisms*

Because of censorship, the visibility to Weibo users of tweets with a certain content may differ from the availability of these tweets in our data set.

Consider two types of content referencing protests, labeled A and B. Type-A content reveals logistics information (e.g., where and when to meet), which lowers the cost of protesting. Type-B content just talks about the occurrence of a protest or expresses sentiment, which supports other mechanisms. The Chinese government censors Type-A content but not Type-B content. Given that data downloading is slower than user reading of tweets, users may read Type-A tweets that are censored before we download them. Hence, the ratio in the downloaded tweets may be a biased measure of the ratio of reader exposure (the share of tweets read by users). Below, we provide a simple formulation to assess this potential bias.

Suppose that there are N Type-A tweets. Among these tweets, let α be the share of those that contain sensitive keywords, which are therefore automatically held in quarantine. A share s of these pass the censorship test and are later published; these published tweets are downloaded by us and also visible to users. The remaining tweets, which amount for a share of $1-\alpha$, do not contain sensitive keywords, and are directly published but later reviewed and perhaps censored. We download a share δ of these tweets. Suppose that u_A users would read these tweets if they were never censored, and that on average a share r of these users read the tweets before they are censored. Thus, we define a measure of user exposure to Type-A content, e_A, to be

$$e_A = N u_A \big(\alpha s + (1-\alpha) r \big).$$

The number of tweets that we download is

$$d_A = N \big(\alpha s + (1-\alpha) \delta \big).$$

Hence, the ratio of user exposure to downloaded tweets is

$$\frac{e_A}{d_A} = u_A \frac{\alpha s + (1-\alpha) r}{\alpha s + (1-\alpha) \delta}.$$

The equivalent ratio for Type-B content, which is not censored, is u_B, because in this case $\alpha = 0$ and $r = \delta = 1$. The relative ratios of user exposure to the number of downloaded tweets for Type-A content and Type-B content depend on the difference between r and δ as well as the discrepancy between u_A and u_B.

Empirical Estimates. To gauge the discrepancy between reader exposure and data downloading (r and δ), we need to know the share of interested users who read the tweets before they are censored, and the share of tweets that are downloaded before censoring. This requires estimates of the speed of (i) censoring, (ii) user reading, and (iii) our downloading of tweets. Let $\widehat{F}_c(t)$ be the empirical distribution of censoring time as measured by Zhu et al. (2013), who provide an accurate measure of the speed of censorship in minutes after a tweet is posted. Let $\widehat{F}_r(t)$ be the empirical distribution of retweet time in our data, which serves as a proxy for the speed of reading tweets. Finally, we assume that the downloading time, $\widehat{F}_d(t)$, is uniformly distributed between 0 and 24 hours. This assumption is reasonable because most of our data are downloaded by lining up users over a 24-hour time window and the posts of each of these users are downloaded at a daily frequency.

The empirical distributions of censoring and retweet speed are plotted in Figure A.2. This figure includes two distributions of retweet speed for (i) all retweets and (ii) retweets of posts referencing protest and strike within two days after a real world event. In the first 10 minutes after a tweet is posted, retweeting (especially for the protest or strike events) is faster than censoring. After this time, the speed of censoring is considerably higher than that of retweeting. Protest and strike tweets are retweeted more quickly than the average tweet. This could be because users are more active in responding to politically sensitive content or because censorship disables the retweeting of sensitive posts after a certain time. For the latter reason, rather than using protest and strikes tweets, we use the retweet distribution based on all tweets, for which the share of censored posts is very small, to calculate the probabilities below.

We simulate data to generate samples of 1000 tweets. For each simulation, we draw independent realizations of the censoring time (c) from the empirical distribution $\widehat{F}_c(t)$, the download time (d) from $\widehat{F}_d(t)$, and 10 realizations of retweet time (r) from $\widehat{F}_r(t)$.[2] The simulated probability of downloading a tweet that is later censored is 0.23, which implies that approximately 77% of the posts that are eventually censored are censored before they are downloaded. Our simulation show that around one-third of the users who would read a tweet have done so by the time of censorship.

To gauge the ratio of r to δ, suppose that there are 100 Type-A tweets susceptible to censorship and 100 uncensored Type-B content. Each tweet would be read by 10 users if censorship did not occur. In this case, the Type-B tweets are read 1000 times and the Type-A tweets are read 330 times. We download 100 of the Type-B tweets and 23 of the Type-A tweets. Hence, the ratio of downloaded Type-A tweets relative to Type-B tweets is 23/100, and the corresponding ratio of user exposure to tweets is 330/1000.

According to our previous formulation, if no posts in category A contain sensitive keywords and are automatically held for review ($\alpha = 0$), then we would underestimate the prevalence of censored Type-A tweets relative to uncensored Type-B tweets by around 30% (1-23/33). In practice, we expect that a significant share of Type-A tweets contain logistic information, which typically includes sensitive keywords, and thus the bias is smaller than when $\alpha = 0$.

The other issue is that the number of interested users may differ across content types ($u_A \neq u_B$). To capture this difference, we compare the numbers of retweets of Type-A and Type-B content, because the average number of retweets per tweet tends to increase in the number of users who read the tweet.

[2]The number 10 is used because the tweets on protest and strike topics are retweeted ten times on average.

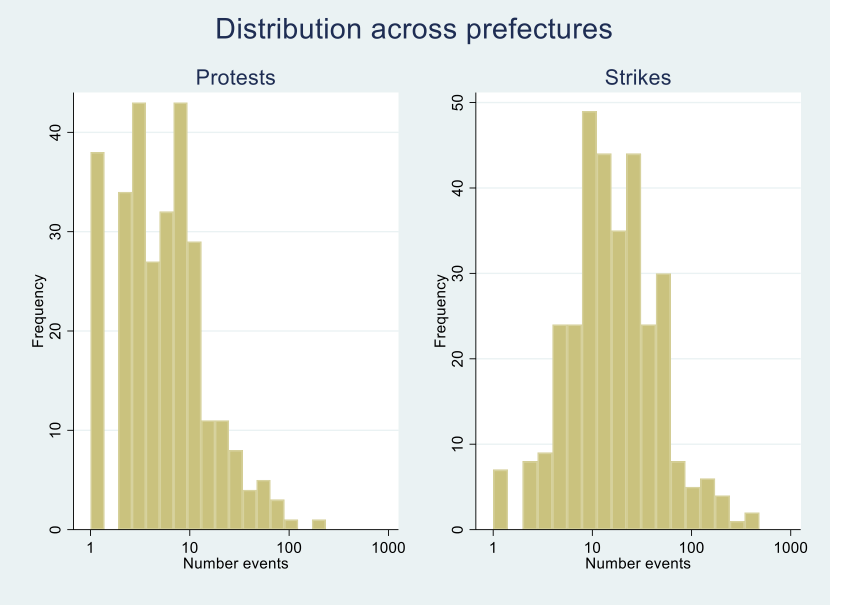

FIGURE A.1.—Distribution of events across cities. *Notes*: The x-axis indicates the accumulated number of events within each city, and the y-axis indicates the frequency of cities. The sample period for protests is from 2006 to 2017, and that for strikes is from 2007 to 2017.

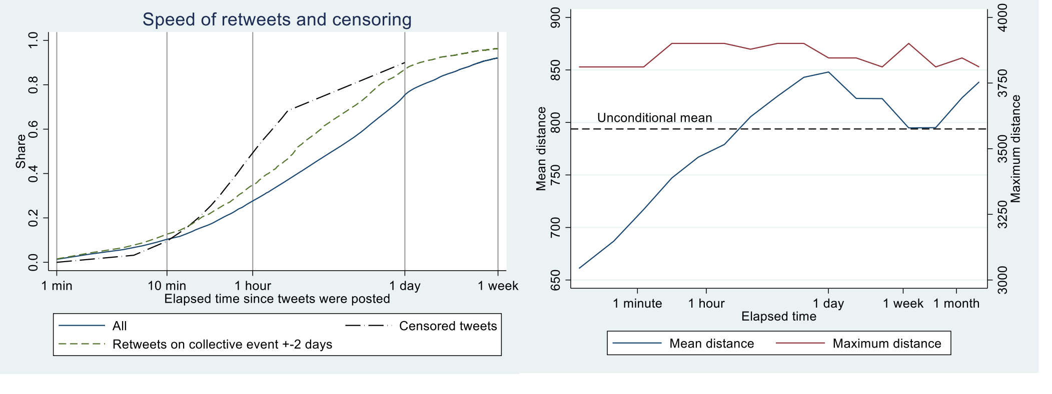

FIGURE A.2.—Speed of retweeting and censorship of posts about protests and strikes.

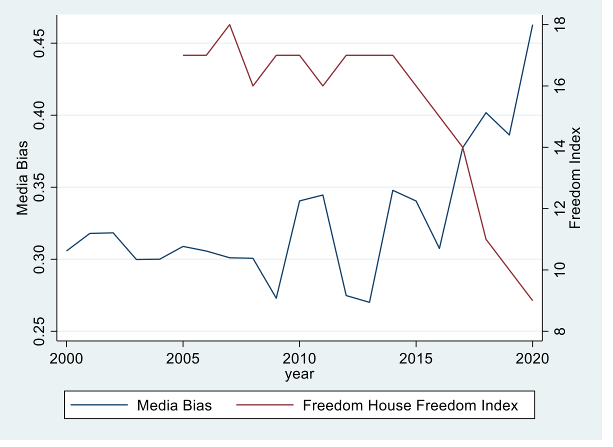

FIGURE A.3.—Media control in China during 2000—2020. *Notes*: "Media Bias" is a measure of pro-government bias of Chinese newspapers based on the method used in Qin et al. (2018). The Freedom House Freedom Index (starting from 2005) is the aggregate score of freedom of expression and media freedom constructed by the Freedom House.

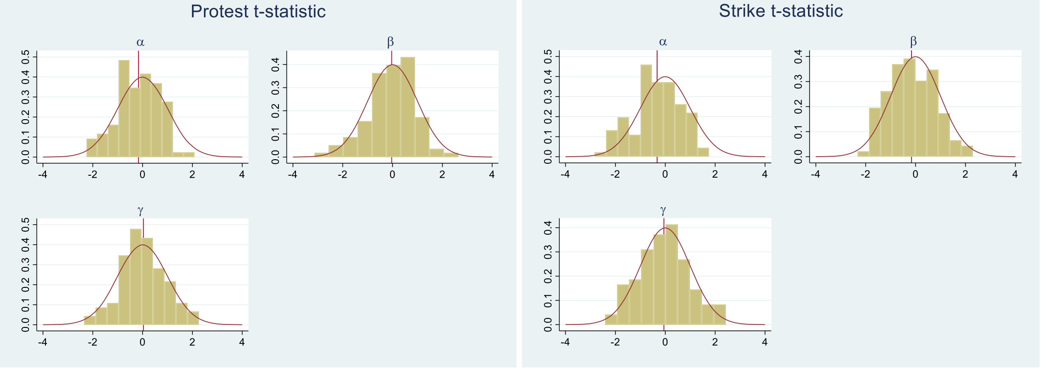

FIGURE A.4.—Monte Carlo simulations DGP, distribution of t-statistics. *Notes*: The graphs plot the distribution of t-statistics of coefficients of interest against the standard normal density. The red vertical line denotes the mean of the standardized t statistics. α, β, γ are coefficients of the three variables in Equation (1): lagged event dummy in city i (y_{it-1}), the time-varying diffusion of information on protests through social media (s_{it-1}), the spread to geographically close cities (d_{it-1}).

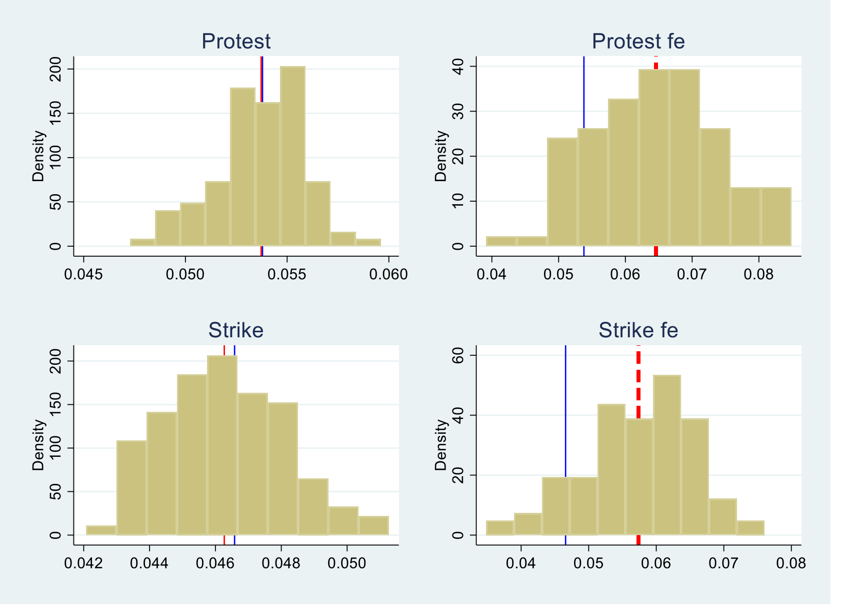

FIGURE A.5.—Monte Carlo simulations of the fixed-effect model. *Notes*: The graphs plot the distribution of β estimates. The blue solid line indicates the coefficient of the DGP. The red dotted line indicates the mean estimated coefficient using the simulated data.

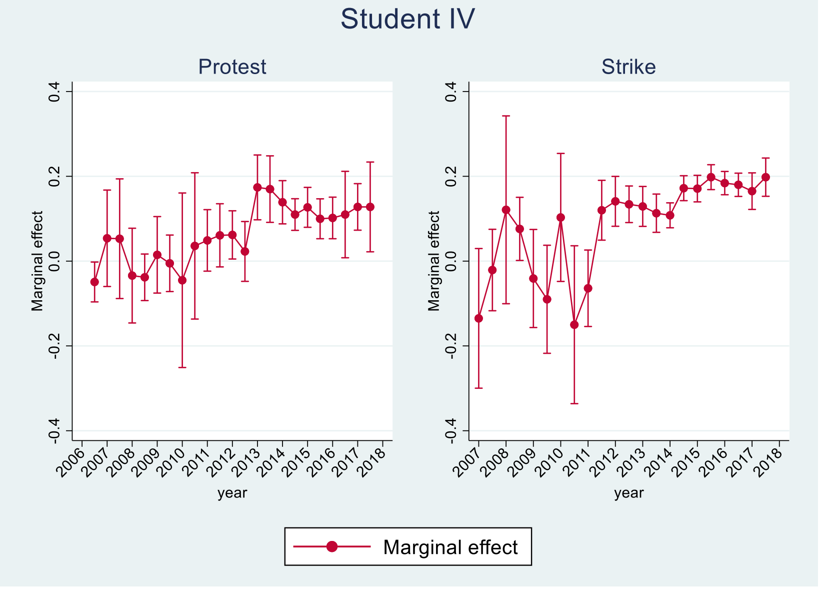

FIGURE A.6.—Student-IV dynamic effects. *Notes*: The graphs plot the estimated biannual coefficients on the time-constant measure $\bar{s}_{it-1}$ (Instrumented by the student-mobility variable z_{it-1}) relative to the pre-period mean. The bars indicate the 95% confidence interval.

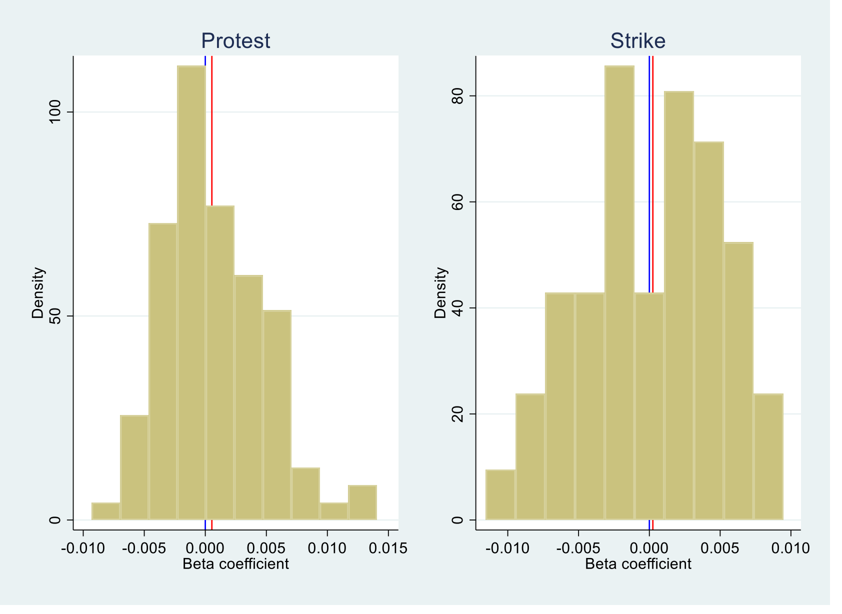

FIGURE A.7.—Monte Carlo simulations with observability driven by Weibo without network spread. *Notes*: The blue (left) line is at the beta-coefficient of the DGP. The red (right) line is at the mean estimated coefficient using the simulated data.

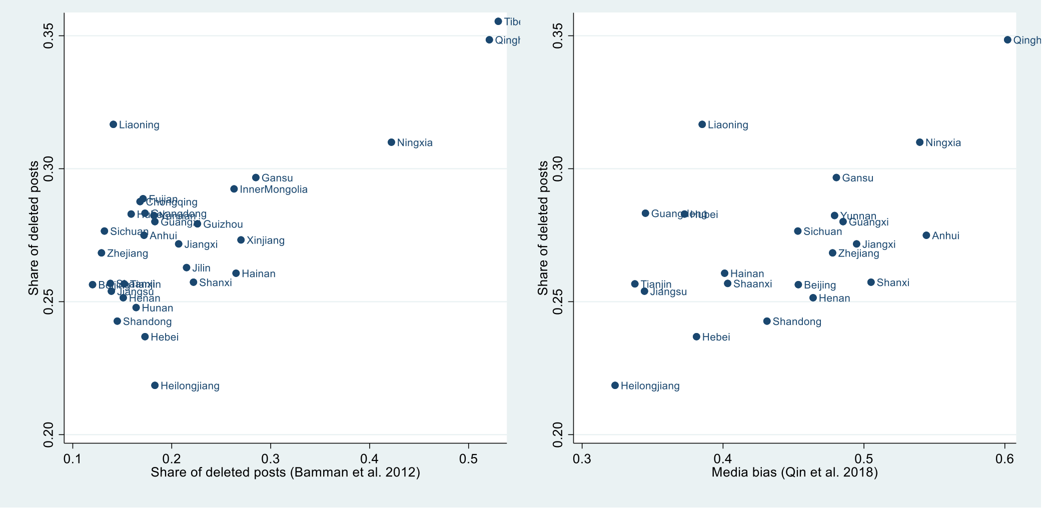

FIGURE A.8.—Local censoring intensity across provinces. *Notes*: The left graph plots the correlation between our measure of censoring intensity (the share of deleted posts) and the measure of the share of deleted posts by Bamman et al. (2012). The right graph plots a similar correlation with the x-variable replaced with a measure of pro-government media bias based on Qin et al. (2018).

TABLE A1

SUMMARY STATISTICS OF THE MAIN VARIABLES.

	Protest				Strike			
	Mean	sd	min	max	Mean	sd	min	max
Event dummy	0.002	0.039	0	1	0.002	0.045	0	1
#events 1–2 days prior (retweet weighted)	0.016	0.04	0	0.69	0.02	0.043	0	0.675
#events 1–2 days prior	0.003	0.056	0	2	0.004	0.065	0	2
#events 1–2 days prior (distance weighted)	0.003	0.007	0	0.145	0.004	0.008	0	0.233
Total # retweets by users in focal city in the last 182 days	1.026	1.019	0	3.495	1.021	0.948	0	3.303
log (#posts per capita +1)	0.06	0.185	0	2.646	0.059	0.18	0	2.646
log(Population)	5.965	0.635	3.767	8.119	5.874	0.684	2.871	8.119
log(GDP)	15.147	1.204	11.989	19.179	15.093	1.184	12.031	19.179
log(Agriculture share of GDP)	1.484	1.068	−3.219	4.078	1.455	1.076	−3.219	4.071
log(Industrial share of GDP)	3.898	0.264	2.123	4.504	3.903	0.277	2.161	4.504
log(Tertiary share of GDP)	3.703	0.266	2.153	4.454	3.693	0.276	2.153	4.454
log(#Cellphone users, 10,000)	5.422	0.846	2.509	8.124	5.362	0.851	2.786	8.124
#cities	248				282			
#observations	670,996				713,702			

TABLE A2

BETWEEN-CITY MOBILITY OF COLLEGE STUDENTS AND SOCIAL MEDIA CONNECTIONS.

Variables	(1) f_{ij}
$\text{Students}_{ij}^{0509}$	0.083
	(0.013)
Inverse geographical distance	39.722
	(4.223)
Observations	86,308
R-squared	0.938
Fixed effects	Prov pair + city-prov pair

Note: The dependent variable is log(1+# of retweets from city i of tweets from city j). $\text{Students}_{ij}^{0509}$ is log(1+mean # students from city i who move to city j in 2005 or in 2009). The regression includes origin and destination city-fixed effects, as well as fixed effects for province pairs and destination city-province pairs. Standard errors (in parentheses) are clustered by origin city and destination city.

TABLE A3

STUDENT-MOBILITY IV: FIRST STAGE.

Variables	Protest $\bar{s}_{it-1}$	Strike $\bar{s}_{it-1}$
Students$_{ij}^{0509}$ *x period 0*	1.196	1.249
	(0.081)	(0.078)
Students$_{ij}^{0509}$ *x period 1*	1.414	1.422
	(0.078)	(0.073)
Students$_{ij}^{0509}$ *x period 2*	1.246	1.494
	(0.075)	(0.081)
Observations	999,472	1,092,477
R-squared	0.980	0.985

Note: The dependent variable is $\bar{s}_{it-1}$, the lagged events in other cities, weighted by social media connections. The unit of observation is city by date. The regression includes lagged events, y_{it-1}, and distance-weighted lagged events, d_{it-1}, interacted with period fixed effects, city-by-period, and date-fixed effects. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors (in parentheses) are two-way clustered by date and city.

TABLE A4

TESTING FOR MECHANICAL ZERO EFFECTS AT LOW EVENT FREQUENCY.

Variables	Mean
β^0	0.115
β^1	0.127
sd(β^0)	0.044
sd(β^1)	0.021
Tests of $\beta^1 > \beta^0$ at 5% significance	0.080

Note: Data are generated so that the simulated event frequencies equal observed event frequencies for the pre-Weibo and post-Weibo periods, while the DGP-beta is the same for both periods. The model of Table II is then estimated on the simulated data.

TABLE A5

TIME-VARYING MEASURE OF CONNECTIONS (LOG MODEL).

Variables	(1) Protest	(2) Protest	(3) Protest	(4) Strike	(5) Strike	(6) Strike
Social-media spread (β)	0.055	0.054	0.047	0.039	0.037	0.024
	(0.013)	(0.012)	(0.013)	(0.008)	(0.007)	(0.008)
Geo-distance spread (γ)	−0.008	−0.007		0.031	0.030	
	(0.008)	(0.008)		(0.013)	(0.012)	
Number events 1–2 days prior	0.009	0.009		0.017	0.017	
	(0.004)	(0.004)		(0.004)	(0.004)	
Total number retweets	0.000	−0.000	0.000	−0.001	−0.001	−0.001
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Weibo posts	0.006	0.006	0.005	0.011	0.010	0.009
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)
Observations	670,996	670,996	670,996	713,702	713,702	713,702
R-squared	0.016	0.016	0.219	0.027	0.027	0.239
Controls	No	Yes	Yes	No	Yes	Yes
QPtest	0.07	0.15		0.27	0.32	

Note: Results are obtained from a linear regression of an event dummy. The unit of observation is city by date. The estimated model is

$$y_{it} = \alpha h(y_{it-1}) + \beta h(s_{it-1}) + \gamma h(d_{it-1}) + \beta_0 w_{it} + \theta' x_{it} + \alpha_{ip} + \delta_t + \varepsilon_{it},$$

where $h(x) = \ln(5x + 1)$. Controls are population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell phone users. Columns (3) and (6) allow for arbitrary time-invariant heterogeneity in the spread across city pairs. In this specification, the variables capturing geo-distance spread and the number of events 1–2 days prior are collinear with the fixed effects and are dropped. The QP statistic reports the p-value of the test for serial correlation in the fixed-effects model. Standard errors (in parentheses) are two-way clustered by date and city.

TABLE A6

TIME-CONSTANT MEASURE OF CONNECTIONS (LOG MODEL).

Variables		(1) Protest	(2) Protest	(3) Strike	(4) Strike
Social-media spread, s_{it-1}	β^0	0.036	0.033	0.023	0.028
		(0.018)	(0.020)	(0.039)	(0.043)
	β^1	0.214	0.246	0.178	0.190
		(0.055)	(0.058)	(0.042)	(0.046)
	β^2	0.100	0.109	0.237	0.249
		(0.027)	(0.028)	(0.032)	(0.035)
Geo-distance spread, d_{it-1}	γ^0	−0.003	−0.010	0.113	0.114
		(0.018)	(0.018)	(0.066)	(0.067)
	γ^1	−0.028	−0.026	0.116	0.113
		(0.038)	(0.038)	(0.059)	(0.061)
	γ^2	0.026	0.024	0.030	0.031
		(0.026)	(0.027)	(0.035)	(0.034)
Observations		1,140,224	1,028,085	1,224,880	1,119,858
R-squared		0.018	0.019	0.048	0.049
Controls		No	Yes	No	Yes
P-value: $\beta^1 = \beta^0$		0.002	0.001	0.006	0.008
P-value: $\beta^2 = \beta^0$		0.048	0.032	0.000	0.000

Note: Results are obtained from a linear regression on an event dummy variable. The unit of observation is city by date. The estimated model is

$$y_{it} = \alpha^p h(y_{it-1}) + \beta^p h(\bar{s}_{it-1}) + \gamma^p h(d_{it-1}) + \beta_0 w_{it} + \theta' x_{it} + \alpha_{ip} + \delta_t + \varepsilon_{it},$$

where $h(x) = \ln(5x + 1)$. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors (in parentheses) are two-way clustered by date and city.

TABLE A7

EVENT SPREAD ACROSS CITIES: PROBIT MODEL.

Variables	(1) Protest	(2) Protest	(3) Strike	(4) Strike
Social-media spread (β)	0.618	0.568	0.558	0.572
	(0.271)	(0.254)	(0.179)	(0.181)
Geo-distance spread (γ)	1.727	1.927	2.129	2.103
	(1.433)	(1.392)	(0.643)	(0.653)
Number events 1-2 days prior	0.178	0.168	0.317	0.319
	(0.079)	(0.077)	(0.049)	(0.048)
Total number retweets	−0.032	−0.042	0.194	0.203
	(0.046)	(0.044)	(0.036)	(0.036)
Weibo posts	0.238	0.232	−0.009	−0.008
	(0.058)	(0.060)	(0.034)	(0.038)
Observations	564,378	564,378	581,509	581,509
Controls	No	Yes	No	Yes

Note: Results are obtained from a probit regression on an event dummy. The unit of observation is city by date. The regression includes city-fixed effects and a quadratic function of the time trend in date. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors (in parentheses) are two-way clustered by date and city.

TABLE A8A

THREE-PERIOD ANALYSIS OF PROTESTS, ADDITIONAL CONTROLS.

Variables		(1) Protest	(2) Protest	(3) Protest	(4) Protest	(5) Protest	(6) Protest	(7) Protest
Social-media spread, s_{it-1}	β^0	0.043	0.029	0.033	0.033	0.031	0.028	0.044
		(0.021)	(0.018)	(0.019)	(0.016)	(0.017)	(0.018)	(0.020)
	β^1	0.204	0.168	0.204	0.179	0.175	0.173	0.189
		(0.061)	(0.045)	(0.056)	(0.046)	(0.045)	(0.045)	(0.065)
	β^2	0.068	0.086	0.090	0.083	0.085	0.086	0.082
		(0.028)	(0.022)	(0.024)	(0.021)	(0.021)	(0.022)	(0.026)
Geo-distance spread, d_{it-1}	γ^0	−0.014	−0.012	−0.013	−0.015	−0.011	−0.012	−0.014
		(0.018)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
	γ^1	−0.034	−0.024	−0.033	−0.030	−0.019	−0.026	−0.027
		(0.037)	(0.035)	(0.037)	(0.037)	(0.034)	(0.036)	(0.038)
	γ^2	0.024	0.020	0.019	0.020	0.022	0.019	0.021
		(0.025)	(0.024)	(0.024)	(0.024)	(0.023)	(0.024)	(0.024)
Observations		1,022,162	1,022,162	1,022,162	1,022,162	1,022,162	1,022,162	1,022,162
R-squared		0.020	0.020	0.020	0.020	0.020	0.020	0.020
Interacted controls		All	Population	GDP	Agriculture share	Industrial share	Tertiary Share	Cell-phone
P-value: $\beta^1 = \beta^0$		0.016	0.004	0.004	0.003	0.003	0.003	0.034
P-value: $\beta^2 = \beta^0$		0.544	0.059	0.089	0.096	0.074	0.059	0.302

Note: Results are from a linear regression on an event dummy variable. The unit of observation is city by date. The regression includes lagged events, y_{it-1} interacted with period-fixed effect, city-by-period, and date-fixed effects. Default controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell phone users. Interacted controls are constructed as $\sum_j w_i w_j y_{jt-1}$, where the weights w_i are population, GDP, etc., for city i. Standard errors (in parentheses) are two-way clustered by date and city.

TABLE A8B

THREE-PERIOD ANALYSIS OF STRIKES, ADDITIONAL CONTROLS.

Variables		(1) Strike	(2) Strike	(3) Strike	(4) Strike	(5) Strike	(6) Strike	(7) Strike
Social-media spread, s_{it-1}	β^0	-0.025	0.034	0.027	0.025	0.031	0.028	0.023
		(0.038)	(0.039)	(0.036)	(0.032)	(0.033)	(0.033)	(0.035)
	β^1	0.023	0.151	0.129	0.126	0.135	0.133	0.092
		(0.038)	(0.035)	(0.030)	(0.026)	(0.029)	(0.029)	(0.031)
	β^2	0.074	0.109	0.133	0.125	0.124	0.125	0.087
		(0.026)	(0.017)	(0.022)	(0.017)	(0.017)	(0.017)	(0.021)
Geo-distance spread, d_{it-1}	γ^0	0.117	0.100	0.102	0.104	0.102	0.102	0.103
		(0.071)	(0.064)	(0.066)	(0.066)	(0.065)	(0.065)	(0.066)
	γ^1	0.112	0.093	0.100	0.102	0.100	0.098	0.106
		(0.052)	(0.051)	(0.054)	(0.056)	(0.054)	(0.054)	(0.055)
	γ^2	0.050	0.043	0.048	0.047	0.054	0.047	0.048
		(0.032)	(0.032)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
Observations		1,113,920	1,113,920	1,113,920	1,113,920	1,113,920	1,113,920	1,113,920
R-squared		0.050	0.050	0.050	0.050	0.050	0.050	0.050
Interacted controls		All	Population	GDP	Agriculture share	Industrial share	Tertiary Share	Cell-phone
P-value: $\beta^1 = \beta^0$		0.263	0.025	0.022	0.005	0.010	0.011	0.113
P-value: $\beta^2 = \beta^0$		0.020	0.055	0.002	0.001	0.004	0.002	0.091

Note: Results are from a linear regression on an event dummy variable. The unit of observation is city by date. The regression includes lagged events, y_{it-1} interacted with period-fixed effect, city-by-period, and date-fixed effects. Default controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Interacted controls are constructed as $\sum_j w_i w_j y_{jt-1}$, where the weights w_i are population, GDP, etc., for city i. Standard errors (in parentheses) are two-way clustered by date and city.

TABLE A8C

THREE-PERIOD ANALYSIS OF PROTESTS, ADDITIONAL CONTROLS (LOG MODEL).

Variables		(1) Protest	(2) Protest	(3) Protest	(4) Protest	(5) Protest	(6) Protest	(7) Protest
Social-media spread, s_{it-1}	β^0	0.045	0.036	0.036	0.035	0.036	0.036	0.037
		(0.024)	(0.020)	(0.020)	(0.019)	(0.020)	(0.020)	(0.020)
	β^1	0.268	0.244	0.244	0.249	0.245	0.244	0.245
		(0.069)	(0.057)	(0.057)	(0.060)	(0.057)	(0.057)	(0.057)
	β^2	0.109	0.107	0.107	0.102	0.107	0.107	0.108
		(0.030)	(0.028)	(0.028)	(0.027)	(0.028)	(0.028)	(0.028)
Geo-distance spread, d_{it-1}	γ^0	−0.011	−0.011	−0.011	−0.011	−0.011	−0.011	−0.011
		(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
	γ^1	−0.034	−0.029	−0.029	−0.031	−0.029	−0.029	−0.030
		(0.039)	(0.037)	(0.037)	(0.038)	(0.038)	(0.037)	(0.037)
	γ^2	0.023	0.023	0.023	0.025	0.023	0.023	0.023
		(0.026)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Observations		1,022,162	1,022,162	1,022,162	1,022,162	1,022,162	1,022,162	1,022,162
R-squared		0.019	0.019	0.019	0.019	0.019	0.019	0.019
Interacted controls		All	Population	GDP	Agriculture share	Industrial share	Tertiary Share	Cell-phone
P-value: $\beta^1 = \beta^0$		0.003	0.001	0.001	0.001	0.001	0.001	0.001
P-value: $\beta^2 = \beta^0$		0.157	0.051	0.051	0.059	0.052	0.051	0.055

Note: Results are from a linear regression on an event dummy variable. The unit of observation is city by date. The regression includes lagged events, y_{it-1} interacted with period-fixed effect, city-by-period, and date fixed effects. Default controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Interacted controls are constructed as $\sum_j w_i w_j y_{jt-1}$, where the weights w_i are population, GDP, etc., for city i. Standard errors (in parentheses) are two-way clustered by date and city.

TABLE A8D

THREE-PERIOD ANALYSIS OF STRIKES, ADDITIONAL CONTROLS (LOG MODEL).

Variables		(1) Strike	(2) Strike	(3) Strike	(4) Strike	(5) Strike	(6) Strike	(7) Strike
Social-media spread, s_{it-1}	β^0	-0.017	0.019	0.018	0.026	0.020	0.019	0.015
		(0.058)	(0.045)	(0.044)	(0.039)	(0.044)	(0.044)	(0.045)
	β^1	0.167	0.183	0.182	0.179	0.183	0.181	0.181
		(0.041)	(0.042)	(0.042)	(0.041)	(0.042)	(0.042)	(0.042)
	β^2	0.258	0.248	0.248	0.247	0.249	0.249	0.249
		(0.036)	(0.035)	(0.034)	(0.034)	(0.035)	(0.035)	(0.035)
Geo-distance spread, d_{it-1}	γ^0	0.120	0.112	0.112	0.110	0.111	0.112	0.113
		(0.074)	(0.068)	(0.068)	(0.068)	(0.067)	(0.068)	(0.068)
	γ^1	0.110	0.106	0.106	0.107	0.105	0.106	0.106
		(0.058)	(0.057)	(0.057)	(0.058)	(0.057)	(0.057)	(0.057)
	γ^2	0.030	0.031	0.031	0.032	0.031	0.031	0.031
		(0.035)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
Observations		1,113,920	1,113,920	1,113,920	1,113,920	1,113,920	1,113,920	1,113,920
R-squared		0.049	0.049	0.049	0.049	0.049	0.049	0.049
Interacted controls		All	Population	GDP	Agriculture share	Industrial share	Tertiary Share	Cell-phone
P-value: $\beta^1 = \beta^0$		0.007	0.006	0.005	0.002	0.005	0.006	0.006
P-value: $\beta^2 = \beta^0$		0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: Results are from a linear regression on an event dummy variable. The unit of observation is city by date. The regression includes lagged events, y_{it-1} interacted with period-fixed effect, city-by-period, and date-fixed effects. Default controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Interacted controls are constructed as $\sum_j w_i w_j y_{jt-1}$, where the weights w_i are population, GDP, etc., for city i. Standard errors (in parentheses) are two-way clustered by date and city.

TABLE A9

CORRELATION BETWEEN RETWEETING AND CENSORING INTENSITY.

Variables	(1) f_{ij}	(2) f_{ij}	(3) f_{ij}
Share deleted i	0.026	0.032	
	(0.030)	(0.035)	
Share deleted j	−0.000	−0.002	
	(0.016)	(0.018)	
Share deleted i x Weibo penetration i			0.209
			(0.210)
Share deleted j x Weibo penetration j			−0.116
			(0.100)
Observations	99,048	73,015	73,015
R-squared	0.750	0.803	0.803
City controls	No	Yes	Yes

Note: Results are obtained from a linear regression of f_{ij} on the share of deleted Weibo posts in a city. All variables are standardized. The unit of observation is city i by city j. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors, clustered by city i and city j, are reported in parentheses.

TABLE A10

TIME-VARYING MEASURE OF CONNECTIONS INTERACTED WITH CENSORING INTENSITY.

Variables	(1) Protest	(2) Protest	(3) Strike	(4) Strike
Social-media spread (β)	0.163	0.162	0.120	0.116
	(0.049)	(0.049)	(0.025)	(0.024)
Social-media spread (β) $\times$ *share deleted posts*	0.024	0.026	0.009	0.015
	(0.044)	(0.044)	(0.026)	(0.029)
Geo-distance spread (γ)	−0.031	−0.031	0.142	0.137
	(0.036)	(0.036)	(0.058)	(0.054)
Number events 1–2 days prior	0.015	0.015	0.030	0.030
	(0.007)	(0.007)	(0.007)	(0.006)
Total number retweets	0.002	0.002	0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.001)
Weibo posts	0.006	0.006	0.010	0.009
	(0.002)	(0.002)	(0.004)	(0.003)
Observations	670,996	670,996	713,702	713,702
R-squared	0.017	0.017	0.027	0.027
Controls	No	Yes	No	Yes

Note: Results are obtained from a linear regression of an event dummy. The unit of observation is city by date. The regression includes city-by-period and date-fixed effects. Controls include population, GDP, shares of the industrial and tertiary sectors in GDP, and the number of cell-phone users. Standard errors (in parentheses) are two-way clustered by date and city.

REFERENCES

BAMMAN, DAVID, BRENDAN O'CONNOR, AND NOAH SMITH (2012): "Censorship and Deletion Practices in Chinese Social Media," First Monday. [0014]

CATTANEO, MATIAS, RICHARD CRUMP, MAX FARRELL, AND YINGJIE FENG (2024): "On Binscatter," *American Economic Review*, 114 (5), 1488–1514. [0005]

MOFFITT, ROBERT A. (2001): "Policy Interventions, Low-Level Equilibria, and Social Interactions," *Social Dynamics*, 4.45-82, 6–17. [0001]

QIN, BEI, DAVID STRÖMBERG, AND YANHUI WU (2018): "Media Bias in China," *American Economic Review*, 108 (9), 2442–2476. [0012,0014]

ZHU, TAO, DAVID PHIPPS, ADAM PRIDGEN, JEDIDIAH CRANDALL, AND DAN WALLACH (2013): "The Velocity of Censorship: High-Fidelity Detection of Microblog Post Deletions," in *USENIX Security Symposium*, 227–240. [0010]

Co-editor Dave Donaldson handled this manuscript.

Manuscript received 7 September, 2021; final version accepted 23 August, 2024; available online 10 September, 2024.