These replication materials reproduce the results of the paper “Authoritarian Elasticity: How Autocracies May Effectively Mobilize for Crisis Management.” The raw dataset contains several million official news articles collected from different levels of Chinese governments and is therefore extremely large. We provide the cleaned datasets used in the analysis, which include counts of different types of articles as well as the full text of posts from local grid centers (filtered).

Code:

1_data_preparation: This script converts the raw data into the datasets used in this paper. As the raw datasets exceed 1 TB in size, please contact the authors if you wish to explore them.

2_RDD_First_Outbreak: This script examines the top-down mobilization of the grid system during the initial COVID-19 outbreak, using data from the official WeChat accounts of 27 out of 31 provincial-level Political and Legal Affairs Committees in China. It employs a regression discontinuity design (RDD) centered on the Wuhan lockdown on January 23, 2020.

3_Local_grid_First_Outbreak: ML + Dirichlet regression on local grid-center posts to track functional transformations of the grid system during COVID-19.

4_RDD_Dynamic_Zero_Covid: Uses city- and province-level release data and multiple RDD estimations to test the mobilization of the grid system under the “Dynamic Zero-COVID” policy.

5_RDD_Dynamic_Zero_Covid_robustness: Robustness check of the mobilization of the grid system under the “Dynamic Zero-COVID” policy using data from the Political and Legal Affairs Committees.

Dataset:

nightlight2020: Global 2020 VIIRS nightlight intensity raster used to calculate average light values for each GC account area.

account_data: Information on each GC account, including location, population, GDP, area, and other regional attributes.

coded: Human-labeled GC articles with assigned topic categories (C/S/O).

daily_first_outbreak: A daily dataset of articles from 27 provincial Political and Legal Affairs Committees (shengjizhengfa) WeChat accounts, recording each account’s total and grid-related posts during the initial COVID-19 outbreak (from the beginning of 2019 to the end of 2020).

daily_first_outbreak_robust: The same daily dataset excluding Jiangsu Province's account, used for robustness checks during the initial outbreak.

grid_centers: Original GC posts with full text and related metadata such as ID, account, and date filtered).

Shengjifabu: The dataset is built from provincial Release (shengji fabu) WeChat accounts, recording each account's daily total number of articles and grid-related articles from its establishment to 2023.

Shengjizhengfa: The dataset is built from provincial Political and Legal Affairs Committees (shengjizhengfa) WeChat accounts, recording each account's daily total number of articles and grid-related articles from its establishment to 2023.

Shijifabu: The dataset is built from city-level Release WeChat accounts (Shengjifabu), recording each account's daily total number of articles and grid-related articles from its establishment to 2023.

Shijizhengfa: The dataset is built from city-level Political and Legal Affairs Committees (shijizhengfa) WeChat accounts, recording each account's daily total number of articles and grid-related articles from their establishment to 2023.

sporadic_test: The dataset documents major city-level (provincial-level) COVID-19 outbreaks during China's Dynamic Zero-COVID period, listing for each episode the date of first infection, affected city, outbreak scale, regional and national spread indicators, and associated provinces.

stopwords: A list of common words (e.g., "的", "了", "and") removed during text preprocessing to reduce noise in analysis.

Generated Data:

dynamic_RDD_results: This folder stores all regression-discontinuity estimates for outbreaks during the Dynamic Zero-COVID period, including local, regional, and national mobilization effects, lagged analyses (7-, 14-, and 21-day), and robustness check results.

for_coding: Randomly selected subset of GC articles (about 30%) exported for manual topic labeling (C/S/O).

classified_samples: Combined dataset containing both manually labeled and machine-predicted topic compositions (C, O, S probabilities) for all GC articles, used as the dependent variables in subsequent regression analyses.

cleaned_df: Tokenized and cleaned word list where each row links an article ID to one Chinese token after removing digits, Latin letters, stopwords, empty tokens, and rare or single-character words.

df_matrix_result: Wide TF–IDF document–term matrix where each row represents an article (by ID) and each column a token, with cell values showing the token's TF–IDF weight used for text classification.

final_GC_articles: Filtered set of GC articles that successfully produced TF–IDF features, each containing full text, metadata, and character count (words) used for later modeling.

final_GC_by_account: Summary table counting how many valid articles each official GC account contributed, used for descriptive statistics and appendix reporting.

local_grid: Article-level dataset combining each post's topic composition (O/C/S) with metadata and account-level covariates (e.g., nightlight2020, GDP, population, city level, Hubei indicator, text length, and time interval) for use in Dirichlet regression analysis.

rf_model1: Trained random forest classifier used to predict article topics (C/S/O) based on TF–IDF features.

rf_model2: A feature-restricted random forest retrained using only the top-50 most important tokens from rf_model1, tuned for optimal mtry to reduce dimensionality, limit overfitting, and improve interpretability and efficiency.

topic_coding: Randomly sampled GC articles exported for manual topic labeling (C/S/O).