Codebook for Locked Down, Lashing Out: COVID-19 Effects on Asian Hate Crimes in Italy

Dipoppa, G., Grossman, G., Zonszein, S.

March 22, 2022

Description of variables in ′./data/panel_month_dummies.RData′ and panel_month_dummies.dta

---

Time and space indicators

- **panelvar:** Unique identifier for each municipality in the sample. N=7,922 municipalities.
- **date_monthly:** Unique identifier for each month in the database from January 2007 to March 2020.
- **month:** Year-Month of hate crime.
- **pro_com:** Census 2011 unique municipality identifier.
- **city:** Municipality name (string).
- **province:** Province name (string).
- **region:** Region name (string).
- **month_xxxx-yy-01:** Year(xxxx)-month(yy) dummies.
- **code_province_x:** Province(x) dummies.

Dependent variables

- **n_hc:** Number of hate crimes per municipality and month-year.
- **asian_hc:** Number of hate crimes against Asians per municipality and month-year.
- **hc_pc:** Number of hate crimes in a municipality-month per 10,000 residents.
- **hc_pc_asians:** Main dependent variable used in the analyses. Number of hate crimes against Asians in a municipality-month per 10,000 residents.
- **hc_pc_african:** Number of hate crimes against Africans in a municipality-month per 10,000 residents.

- **hc_pc_asians_nonverbal:** Number of hate crimes against Asians in a municipality-month per 10,000 residents excluding all hate crimes that are not verbal.
- **hc_pc_asians_verbal:** Number of hate crimes against Asians in a municipality-month per 10,000 residents excluding all hate crimes that are purely verbal.
- **hc_pc_asians_violent:** Number of hate crimes against Asians in a municipality-month per 10,000 residents including only instances of physical violence.
- **hc_pc_asians_dropnationalmedia:** Number of hate crimes against Asians in a municipality-month per 10,000 residents excluding all news items reported by national media.
- **hc_pc_asians_sourcenotindividual:** Number of hate crimes against Asians in a municipality-month per 10,000 residents excluding all news items reported by individuals to the NGO.
- **hc_pc_asians_attacker_instit** Number of hate crimes against Asians in a municipality-month per 10,000 residents subsetting to crimes committed by media, politicians or institutions.
- **hc_pc_asians_attacker_person** Number of hate crimes against Asians in a municipality-month per 10,000 residents subsetting to crimes committed by persons or groups.
- **hc_pc_asians_nosocialmedia:** Number of hate crimes against Asians in a municipality-month per 10,000 residents excluding all items which mention the word Facebook, Instagram or Youtube (i.e. crimes possibly happening on social media).

Independent variables

- **covid2:** Dummy equal to 1 for months after January 2020, and 0 otherwise.
- **losing_income_d:** Economic threat treatment indicator. Dummy equal to 1 when a municipality has a share of the population employed in Covid-affected sectors larger than median. Covid-affected sectors include transportation, tourism, and restaurants.
- **workers_losing_income_share:** Alternative economic threat treatment indicator. Share of a municipality's population employed in Covid-affected sectors. Covid-affected sectors include transportation, tourism, and restaurants.
- **losing_income_noturism:** Alternative economic threat treatment indicator. Dummy equal to 1 when a municipality has a share of the population employed in Covid-affected sectors excluding tourism larger than median. Covid-affected sectors include transportation and restaurants.
- **excess_deaths_jan_d:** Health threat treatment indicator. Dummy equal to 1 when the estimated excess deaths in January 2020 are above the median of municipalities.
- **excess_deaths_feb_d:** Alternative health threat treatment indicator. Dummy equal to 1 when the estimated excess deaths in February 2020 are above the median of municipalities.

- **extr_right_mayor:** Dummy equal to 1 for municipalities and months in which a city council is governed by a far-right mayor.
- **extr_right_d:** Dummy equal to 1 for municipalities and months with a vote share for far-right parties in national elections above the median municipality value.
- **covid_econ:** Interaction between covid2 and losing_income_d.
- **covid_unemp:** Interaction between covid2 and losing_income_d.
- **covid_death:** Interaction between covid2 and excess_deaths_jan_d.
- **covid_farright:** Interaction between covid2 and extr_right_mayor.
- **unemp_farright:** Interaction between losing_income_d and extr_right_mayor.
- **covid_unemp_farright:** Triple interaction between covid2, losing_income_d and extr_right_mayor.

Control and other variables

- **timevar** Unique identifier for each month-year in the database from January 2007 to March 2020.
- **timevar_losingincome:** Interaction between timevar and losing_income_d.
- **population:** Municipal level population as from Population Census 2011.
- **foreign_pop:** Municipal level foreign population as from Population Census 2011.
- **educ_diplba:** Share of the municipal population with a bachelor degree or higher titles, from Population Census 2011.
- **unemp:** Unemployment rate in a municipality, from Population Census 2011.
- **unemp_male:** Male unemployment rate in a municipality, from Population Census 2011.
- **unemp_young:** Unemployment rate for young individuals (15-29 years old) in a municipality, from Population Census 2011.
- **emp_unskilled:** Share of unskilled labor in a municipality from Population Census 2011.
- **pop_dens:** Population density, from Population Census 2011.
- **pop15_34:** Population between 15 and 34 years old in a municipality as from Population Census 2011.
- **above75_share:** Population share of the inhabitants of a municipality above 75 years old as from Population Census 2011.

- **share_asian_immigrant:** Population share of immigrants coming from East and Southeast Asia in a municipality as from Population Census 2011.
- **vs_extr_right:** Vote share for far-right parties in national elections by municipality month.

Description of variables in ′./data/panel_month_dummies_all_2020_months.RData′

- Includes all variables in panel_month_dummies.RDta but extends the observations to all months in 2020.

Description of vars in ′./data/2007_2020_hate_crimes_classified.csv

- **Date:** Date of the hate crime.
- **Description:** Description of the hate crime as recorded by Lunaria (in Italian)
- **region:** Classification of the victim's region of origin (e.g. West Africa).
- **region_broad:** Classification of victim's broad region of origin (e.g. Africa).
- **atti:** String containing the categorization of the type of event. Categories include verbal attacks, physical attacks, discrimination and property damage.
- **perpetratori:** String containing the identity of the attacker. Categories include media, institutions, politics and society.
- **immigrant:** Non-empty when description of crime contains terms referring to immigrants, refugees and asylees.
- **fascism:** Non-empty when description of crime contains fascist terms.
- **chinese:** Non-empty when description of crime contains terms related to Chinese people/China.
- **religion_broad:** Classification of the victim's religion.

Description of vars in ′./data/hate_crimes_classified.csv

- Include the same variables as in ′./data/2007_2020_hate_crimes_classified.csv but excludes observations after March 2020.

Description of vars in ′./data/newspaper_articles_sentiment_classification.csv

- **sentence_index:** Sentence index of newspaper article which contains relevant token.
- **token:** Relevant token extracted from article sentence.
- **prob_positive:** Probability that sentiment of sentence containing relevant token is positive.

- **sentiment_negative:** Indicator variable for prob_positive smaller or equal to 0.5.
- **group:** Whether the token relates to African or Chinese individuals.
- **date:** Article's date.

Description of vars in './data/tweets_classified_pretrained_model.csv

- **id:** Tweet unique identifier.
- **date:** Date of the tweet.
- **prob_positive:** Probability that the sentiment of the sentence in tweet containing relevant token is positive.
- **sentence_index:** Sentence index of tweet which contains relevant token.
- **token:** Relevant token extracted from article sentence.
- **group:** Whether the token relates to African or Chinese people.
- **clean_text:** Text of tweet post-text processing.
- **text:** Text of tweet pre-text processing.
- **sentiment_negative:** Indicator variable for prob_positive smaller or equal to 0.5.

Description of vars in './data/municipality_monthly_deaths_2007_2020.csv

- **panelvar:** Municipality unique identifier.
- **month:** Year-monthly unique identifier.
- **deaths:** Number of recorded deaths.

Description of vars in './data/excess_deaths_cdc_methodology.RData

- **panelvar:** Municipality unique identifier.
- **excess_deaths_jan:** Estimated excess deaths in January 2020.
- **excess_deaths_feb:** Estimated excess deaths in February 2020.
- **excess_deaths_jan_d:** Estimated excess deaths in January 2020 are above the median municipality.
- **excess_deaths_feb_d:** Estimated excess deaths in February 2020 are above the median municipality.

Description of vars in './data/unemployment_region.csv

- **region:** String, name of region.
- **unemployment:** Unemployment rate by Italian region in the first two quarters of 2020, from Istat.
- **losing_income_region:** Share of workers in the hospitality industry by region in 2011 as from Census 2011.

Description of vars in './data/Itanes_2018_release01_panel_pre_post.dta

- **caseid:** Unique respondent identifier.
- **dem01:** Gender.
- **dem02:** Age
- **dem05:** Province of residence.
- **dem09:** Education level.
- **dem11:** Employment.
- **img1:** Answer to the question of whether we receive too many immigrants.
- **lrself:** Answer to self-positioning in a left-right spectrum (1-7).
- **voto1:** Self-declared voting.

Description of vars in './data/data_to_merge_with_mobility.csv

- **COD_PROV:** Province unique identifier.
- **excess_deaths_jan_d:** indicator variable for whether estimated excess deaths in January 2020 are above the median municipality.

Description of vars in './data/mobility_median_network.csv

- **COD_PROV:** Province unique identifier.
- All the rest of variables. Median of users' radius of gyration in a province week.

Description of vars in './data/mobility_average_network.csv

- **COD_PROV:** Province unique identifier.
- All the rest of variables. Daily time-series of the average degree of the proximity network.

Description of vars in './data/covid_share_weekly_data.csv

- **date:** Tweet's date.
- **week:** Tweet's week number.
- **share_of_covid:** Share of tweets that contains covid-related keywords aggregated by week.
- **excess_deaths_jan_d:** Indicator variable for whether estimated excess deaths in January 2020 are above the median municipality.
- **losing_income_d:** Indicator variable for whether a municipality has a share of the population employed in Covid-affected sectors larger than the median.

Description of vars in './data/localita11.dta and './data/coord−11.dta

---

- **id:** Unique spatial identifier of municipalities in Census 2011.
- **ID:** Unique spatial identifier of municipalities in Census 2011.
- **x_cc:** Latitude of municipalities in Census 2011.
- **y_cc:** Longitude of municipalities in Census 2011.
- **COD_REG:** Unique identifier of regions in Census 2011.
- **COD_PRO:** Unique identifier of provinces in Census 2011.
- **pro_com:** Unique identifier of municipalities in Census 2011.
- **city:** String, city name.
- **region:** String, region name.