Coverage of Corruption in a Sample of Mexican Newspapers: Codebook

Piero Stanig*

Two datasets are released as replication data for the paper "Regulation of Speech and Media Coverage of Corruption: An Empirical Analysis of the Mexican Press." One is an article-level dataset, that simply reproduces in tabular form the information hand-coded (on paper records) from the content analysis carried out on the hard copies (and, in some specific cases, the electronic archives) of the newspapers. The second is a dataset with observations at the newspaper-day level: this is a proper replication dataset that makes it possible to estimate models like those reported in the paper.

1 Article-level dataset

The article-level dataset is released for the purpose of ensuring third-party evaluation, replication, and extension, of the content analysis, and it should not be analyzed as is. In order to generate the dataset analyzed in the paper, one needs to run the Stata do file `fromarticlestodata.do`. This turns the article-level data into a newspaper-day dataset, which, merged with control variables, becomes the replication dataset described in the next section.

NWSP : name of the newspaper

date : date of publication, as character string (in Day/Month/Year) format.

HEAD_N : progressive number of the article during data collection

HEAD_FULL : short summary of the headline, to make it possible to identify exaclty which article was coded

PAGE : page of the newspaper

SECTION : section of the newspaper.

REMARK : additional notes about the article

ALLEG_TYPEX : with X$\in \{1, 2, 3\}$, type of the allegation, qualitatively coded. Please refer to the paper and the value labels in the Stata dataset.

*Hertie School of Governance, Berlin. stanig@hertie-school.org and piero.stanig@gmail.com. Web: http://sites.google.com/site/pierostanig

Given that an article can report events that fall in more than one category, up to three types per article can be recorded. In addition, further classification is possible, based on the notes reported in **ALLEG_OTHER**.

ALLEG_OTHER : character variable, additional information about the allegation in the article.

LOCAL_SPECIFIC : if equal to 0, it implies that no articles fulfilling the pre-requisites to be counted is present in the newspaper in the sampled date. This is the only record for newspaper-day observations with no articles.

publishing_firm : name of the publishing group. Please refer to the OEM, sintesis and novedades dummies instead.

date_prop : date of publication, in Stata date format.

state : state of publication.

state_short : state of publication, short string variable.

CULPRIT : qualitative coding of the position of the main "culprit" discussed in the article. Coded as politician, bureaucrat (a category that includes street-level civil servants), court actor (e.g., judge, prosecutor), member of a police force (including customs officers), or member of the military.

CULPRIT_LEVEL : qualitative coding of the main "culprit" as part of the federal, state, or municipal government. Some values are missing due to the impossibility to extract this information from the article.

vote_buy : qualitative coding, dummy variable for articles that deal with incidents of vote-buying.

elect_fraud : qualitative coding, dummy variable for articles that deal with alleged electoral fraud.

hr_abuse : qualitative coding, dummy variable for articles that allege violations of human rights by security forces (police and army).

desvio : qualitative coding, dummy variables for articles that allege *desvio*, resource diversion.

not_corruption : dummy for the articles that were not referring to corruption. Refer to the paper for a discussion of the definitions.

OEM : dummy for the newspapers that belong to the OEM publishing group.

sintesis : dummy for the newspapers that belong to the Sintesis group.

novedades : dummy for the newspapers that belong to the Novedades group.

Notice that **only the information about newspaper, date, and characteristics of the culprit, is used in the calculations that lead to the newspaper-day level counts** contained in the replication dataset. The information about the publishing group is used to create the dummies for publishing firms. The remainder of the information (for instance, the allegation type variables) allows for identification of the articles in the original sources. It can potentially be used for further analysis (to create different counts of articles, e.g., restricting attention to human right abuses) but the author of the original dataset is, obviously, not responsible for such uses.

2 Replication dataset

This dataset can be merged with state-level information (e.g., gdp per capita) and used to replicate the results reported in the paper. The dataset is distributed in R data (.Rdata) format and can be easily loaded in `R` for analysis. This is a description of the variables.

NWSP : name of the newspaper.

dow : daily observation identifier.

state : state of publication.

OEM : dummy for newspapers that belong to the OEM publishing group.

novedades : dummy for newspapers that belong to the Novedades publishing group.

sintesis : dummy for newspapers that belong to the Síntesis publishing group.

articles : count of articles about corruption in the state of publication (see the paper for definitions).

politics_articles : count of articles about corruption, involving culprits coded as politicians or bureaucrats.

hilevel : count of articles about high-level corruption, namely involving federal and state-level politicians as culprits.

petty : count of articles about petty corruption, namely involving members of the police forces and municipal-level culprits.

police : count of articles about police misconduct (i.e., those that involve one or more members of the police forces as culprit).

maximum_sanctions : maximum sentence for defamation in effect in 2001. Source: penal code of each state, consulted in the Arthur W. Diamond Law Library at Columbia University.

congressonemonth_away : dummy for newspaper-day observations in the month prior to an election.

congresssixmonths_away : dummy for newspaper-day observations between six and one months prior to an election.

min_homicide : minimum sentence for the crime of *homicidio* in effect in 2001. Source: penal code of each state, consulted in the Arthur W. Diamond Law Library at Columbia University.

max_evasion : maximum sentence for the crime of *evasión de presos* in effect in 2001. Source: penal code of each state, consulted in the Arthur W. Diamond Law Library at Columbia University.

defamation_charged_pc_av : average (1998-2000) number of defamation indictments (*presuntos*) over population. Source: INEGI, Judicial Statistics.

corruption : Corruption score for 2001. Source: Transparencia Mexicana.

corruption.obj : Corruption score, from the estimates reported as *Índice de honestidad y eficiencia en la generación de infraestructura pública* by CEI Consulting & Research.

corruption.tec : Ordinal measure of state capture, from the estimates by Instituto Tecnológico de Estudios Superiores de Monterrey and Centro de Estudios Estratégicos (CEE).