Codebook

This codebook contains the definition of variables used in Steven Liao. "The Effect of Firm Lobbying on High-Skilled Visa Adjudication." *The Journal of Politics*.

Main Datasets

- `lv-imm-fr-99-17-rep.RData`: A report-level dataset used in the analysis "Figure-1.R". The dataset contains 18 variables and 7,297 observations.
  - `report_id`: Report ID from LobbyView (Kim 2018)
  - `year`: Year
  - `report_type`: Report type from LobbyView (e.g., quarter report)
  - `period`: Report period from LobbyView (e.g., first quarter)
  - `client_bvdid`: Client BvD ID
  - `client_name`: Client name
  - `issue`: General issue area code (e.g., "IMM")
  - `rep_amount`: Total lobbying expenses in the report
  - `n_issues`: Total number of general issue area codes included in the report
  - `n_imm`: Total number of "IMM" issue codes included in the report
  - `mean_img_expense_report`: Weighted average immigration lobbying expenses in the report
  - `venue`: Target lobby venue(s)
  - `lob_dhs`: Lobbied the Department of Homeland Security (DHS) on immigration? 1 = Yes, 0 = No
  - `lob_wh`: Lobbied the White House (WH) on immigration? 1 = Yes, 0 = No
  - `lob_eop`: Lobbied the Executive Office of the President (EOP) on immigration? 1 = Yes, 0 = No
  - `lob_dol`: Lobbied the Department of Labor (DOL) on immigration? 1 = Yes, 0 = No
  - `lob_house`: Lobbied the House of Representatives on immigration? 1 = Yes, 0 = No
  - `lob_senate`: Lobbied the Senate on immigration? 1 = Yes, 0 = No
- `lv-imm-fy-99-17-rep.RData` : A firm-year dataset used in the analysis "Figure-3.R". The dataset contains 5 variables and 2,196 observations.
  - `client_bvdid`: Client BvD ID
  - `client_name`: Client name
  - `year`: Year
  - `n_img_rep_year`: Total number of reports filed containing "IMM"

- `est_img_expense_fy`: Total estimated immigration lobbying expenditure

- `data-merge-91-17-orbis-excluded.RData`: An unbalanced firm-year dataset that excludes proprietary Orbis variables.[1] The dataset is the same as `data-merge-91-17.RData` but with Orbis variables excluded. The latter full dataset is documented in the code but not shared for proprietary data protection reasons. The dataset is used in analyses:

  - Figure-4-Appendix-Table-D3-D4-D5.R
  - Figure-6-Appendix-Figure-D1.R
  - DID-analysis.R
  - Appendix-Figure-B1.R
  - Appendix-Table-D1.R.

  The dataset contains 64 variables and 1,243,396 observations.

  - `year`: Year
  - `bvd_id`: Orbis BvD ID
  - `bvd_name`: Firm's name from Orbis
  - `h1b_pet_name`: Firm's name from the author's H-1B petition data
  - `h1b_pet_state`: Firm's state from the author's H-1B petition data
  - `h1b_pet_zcta`: Firm's ZCTA from the author's H-1B petition data
  - `l1_pet_name`: Firm's name from the author's L-1 petition data
  - `l1_pet_state`: Firm's state from the author's L-1 petition data
  - `l1_pet_zcta`: Firm's ZCTA from the author's L-1 petition data
  - `h1b_pet_firm_year_tot`: Total H-1B petitions by the firm in the year
  - `n_img_rep_year`: Total number of lobbying reports containing "IMM" filed by the firm in the year
  - `est_img_expense_fy`: Total estimated immigration lobbying expenditure by the firm in the year
  - `lob_img_2016`: Lobbied on immigration in 2016 (filed at least one report containing "IMM" in 2016)? 1 = Yes, 0 = No
  - `lob_img_2017`: Lobbied on immigration in 2017 (filed at least one report containing "IMM" in 2017)? 1 = Yes, 0 = No
  - `lob_img_2017_aug`: Augmented version of `lob_img_2017`. Filed at least one lobbying report that either included "IMM" as a general issue code or omitted "IMM" but contained some immigration-related keywords ("immigration", "visa", or "H-1B")? 1 = Yes, 0 = No
  - `h1b_deny_rate`: H-1B denial rate. Equals the firm's total denied petitions divided by total petitions (approved, denied, revoked, and administratively closed) in the year

[1] See https://www.bvdinfo.com/en-us/ for Orbis database details.

- post_2017: Year $\geq$ 2017? 1 = Yes, 0 = No
- post_2004: Year $\geq$ 2004? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_2017: A text-refined measure of lobbying. lob_img_2017 = 1 and reports contained keywords "Visa" and either "Skilled" or "H-1B"? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_2017_aug: A text-refined measure of lobbying. lob_img_2017_aug = 1 and reports contained keywords "Visa" and either "Skilled" or "H-1B"? 1 = Yes, 0 = No
- lob_uscis_2017: A venue-refined measure of lobbying. Lobbied the U.S. Citizenship and Immigration Services (USCIS) on immigration in 2017 (filed at least one report containing "IMM" in 2017 that listed the USCIS as a target venue)? 1 = Yes, 0 = No
- lob_dol_2017: A venue-refined measure of lobbying. Lobbied the DOL on immigration in 2017 (filed at least one report containing "IMM" in 2017 that listed the DOL as a target venue)? 1 = Yes, 0 = No
- lob_dhs_2017: A venue-refined measure of lobbying. Lobbied the DHS on immigration in 2017 (filed at least one report containing "IMM" in 2017 that listed the DHS as a target venue)? 1 = Yes, 0 = No
- lob_wh_eop_2017: A venue-refined measure of lobbying. Lobbied the WH/EOP on immigration in 2017 (filed at least one report containing "IMM" in 2017 that listed the the WH/EOP as a target venue)? 1 = Yes, 0 = No
- lob_only_congress_2017: A venue-refined measure of lobbying. Only lobbied the House or the Senate on immigration in 2017 (filed at least one report containing "IMM" in 2017 with all reports only listing the House/Senate as target venues)? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_uscis_2017: A text and venue-refined measure of immigration lobbying in 2017. lob_text_hs_h1b_visa_2017 = 1 & lob_uscis_2017 = 1? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_uscis_2017_aug: A text and venue-refined measure of immigration lobbying in 2017. lob_text_hs_h1b_visa_2017_aug = 1 & lob_uscis_2017 = 1? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_dol_2017: A text and venue-refined measure of immigration lobbying in 2017. lob_text_hs_h1b_visa_2017 = 1 & lob_dol_2017 = 1? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_dol_2017_aug: A text and venue-refined measure of immigration lobbying in 2017. lob_text_hs_h1b_visa_2017_aug = 1 & lob_dol_2017 = 1? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_dhs_2017: A text and venue-refined measure of immigration lobbying in 2017. lob_text_hs_h1b_visa_2017 = 1 & lob_dhs_2017 = 1? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_dhs_2017_aug: A text and venue-refined measure of immigration lobbying in 2017. lob_text_hs_h1b_visa_2017_aug = 1 & lob_dhs_2017 = 1? 1 = Yes, 0 = No

- lob_text_hs_h1b_visa_wh_eop_2017: A text and venue-refined measure of immigration lobbying in 2017. lob_text_hs_h1b_visa_2017 = 1 & lob_wh_eop_2017 = 1? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_wh_eop_2017_aug: A text and venue-refined measure of immigration lobbying in 2017. lob_text_hs_h1b_visa_2017_aug = 1 & lob_wh_eop_2017 = 1? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_only_congress_2017: A text and venue-refined measure of immigration lobbying in 2017. lob_text_hs_h1b_visa_2017 = 1 & lob_only_congress_2017 = 1? 1 = Yes, 0 = No
- lob_text_hs_h1b_visa_only_congress_2017_aug: A text and venue-refined measure of immigration lobbying in 2017. lob_text_hs_h1b_visa_2017_aug = 1 & lob_only_congress_2017 = 1? 1 = Yes, 0 = No
- lob_only_tob_2017: Filed at least one report in 2017 containing "TOB" (Tobacco) but zero reports containing "IMM"? 1 = Yes, 0 = No
- lob_only_bev_2017: Filed at least one report in 2017 containing "BEV" (Beverage) but zero reports containing "IMM"? 1 = Yes, 0 = No
- lob_only_cdt_2017: Filed at least one report in 2017 containing "CDT" (Commodities) but zero reports containing "IMM"? 1 = Yes, 0 = No
- year_1992: Year = 1992? 1 = Yes, 0 = No
- year_1993: Year = 1993? 1 = Yes, 0 = No
- year_1994: Year = 1994? 1 = Yes, 0 = No
- year_1995: Year = 1995? 1 = Yes, 0 = No
- year_1996: Year = 1996? 1 = Yes, 0 = No
- year_1997: Year = 1997? 1 = Yes, 0 = No
- year_1998: Year = 1998? 1 = Yes, 0 = No
- year_1999: Year = 1999? 1 = Yes, 0 = No
- year_2000: Year = 2000? 1 = Yes, 0 = No
- year_2001: Year = 2001? 1 = Yes, 0 = No
- year_2002: Year = 2002? 1 = Yes, 0 = No
- year_2003: Year = 2003? 1 = Yes, 0 = No
- year_2004: Year = 2004? 1 = Yes, 0 = No
- year_2005: Year = 2005? 1 = Yes, 0 = No
- year_2006: Year = 2006? 1 = Yes, 0 = No
- year_2007: Year = 2007? 1 = Yes, 0 = No
- year_2008: Year = 2008? 1 = Yes, 0 = No
- year_2009: Year = 2009? 1 = Yes, 0 = No
- year_2010: Year = 2010? 1 = Yes, 0 = No

  - `year_2011`: Year = 2011? 1 = Yes, 0 = No
  - `year_2012`: Year = 2012? 1 = Yes, 0 = No
  - `year_2013`: Year = 2013? 1 = Yes, 0 = No
  - `year_2014`: Year = 2014? 1 = Yes, 0 = No
  - `year_2015`: Year = 2015? 1 = Yes, 0 = No
  - `year_2016`: Year = 2016? 1 = Yes, 0 = No
  - `year_2017`: Year = 2017? 1 = Yes, 0 = No

  The Orbis variables below are excluded from the publicly shared dataset:

  - `naics_core`: Firm's 4-digit NAICS industry code
  - `naics_core_descrip`: Firm's 4-digit NAICS industry description
  - `size`: Firm's size (small, medium, large, very large)
  - `public`: Publicly listed? 1 = Yes, 0 = No
  - `sales`: Sales (thousand USD)
  - `n_employ`: Total number of employees

- `lv-imm-fr-2008-2017-checked-orbis-excluded.RData`: A manually checked report-level dataset (2008–2017) that excludes proprietary Orbis variables and data. It shows what the text analysis input data looks like in analysis "Figure-5-C2-C3-C4.R" but without Orbis variables. The dataset contains 13 variables and 5,705 observations.

  - `client_bvdid`: Client BvD ID
  - `year`: Year
  - `client_name`: Client name
  - `client_legal_name`: Client legal name from LobbyView
  - `report_id`: Report ID from LobbyView
  - `period`: Report period from LobbyView
  - `date_filed`: Report file date
  - `text`: Specific lobbying issues described under the general issue area code
  - `venue`: Target lobby venue(s)
  - `rep_amount`: Total lobbying expenses in the report
  - `n_issues`: Total number of general issue area codes included in the report
  - `n_imm`: Total number of "IMM" issue codes included in the report
  - `mean_img_expense_report`: Weighted average immigration lobbying expenses in the report

- `lv-imm-fr-1999-2007-checked-orbis-excluded.RData`: A manually checked lobbying report-level dataset (1999–2007) that excludes proprietary Orbis variables and data. This dataset is discussed in Section 4 "Data and Measures" of the manuscript. The dataset contains 13 variables and 1,565 observations.

  - client_bvdid: Client BvD ID
  - year: Year
  - client_name: Client name
  - client_legal_name: Client legal name from LobbyView
  - report_id: Report ID from LobbyView
  - period: Report period from LobbyView
  - date_filed: Report file date
  - text: Specific lobbying issues described under the general issue area code
  - venue: Target lobby venue(s)
  - rep_amount: Total lobbying expenses in the report
  - n_issues: Total number of general issue area codes included in the report
  - n_imm: Total number of "IMM" issue codes included in the report
  - mean_img_expense_report: Weighted average immigration lobbying expenses in the report

Supplementary Datasets

- lv-img-reports-99-17-raw.RData: A report-issue-level raw dataset from 1999 to 2017 focusing on reports containing IMM issue codes. The query data were downloaded from LobbyView's website in October 2019 and are used in "prep-lobbyview-data.R". The dataset contains 17 variables and 166,244 observations.
  - report_id: Report ID
  - report_type: Report type
  - client_bvdid: Client BvD ID
  - client_gvkey: Client GVKEY
  - client_name: Client name
  - client_legal_name: Client legal name
  - client_naics: Client NAICS
  - year: Report year
  - period: Report period
  - date_filed: Report filing date
  - issue: General issue area code
  - text: Specific lobbying issues described under the general issue area code
  - rep_amount: Total lobbying expenses in the report
  - venue: Target lobby venue(s)
  - lobbyists: Names of lobbyists

  - report_has_amendments: Report has amendments? TRUE = Yes, FALSE = No
  - report_is_latest_amend: Report is the latest amendment? TRUE = Yes, FALSE = No

- lv-check-v1-venue-text-2000-2007.csv: A report-level dataset containing some manually checked information on immigration lobbying venues and text (2000–2007). The dataset is used in "prep-lobbyview-data.R" and contains 8 variables and 416 observations.
  - report_id: Report ID
  - report_year: Report year
  - report_type: Report type
  - client_name: Client name
  - report_amount: Total lobbying expenses in the report
  - report_issue: General issue area code
  - report_specific_issues: Specific lobbying issues described under the general issue area code
  - report_venue: Target lobby venue(s)

- lv-check-v2-venue-text-1999-2017.csv: A report-level dataset containing some manually checked information on immigration lobby venues and text (1999–2017). The dataset is used in "prep-lobbyview-data.R" and contains 10 variables and 1,537 observations.
  - report_id: Report ID
  - client_name: Client name
  - client_legal_name: Client legal name
  - year: Report year
  - period: Report period
  - date_filed: Report filing date
  - issue: General issue area code
  - text: Specific lobbying issues described under the general issue area code
  - venue: Target lobby venue(s)
  - drop: Incorrect LobbyView information about report (e.g., does not exist, unrelated to IMM, no lobbying activity, or client is not a firm)? 1 = Yes, 0 = No

- lv-check-v3-amount-1999-2007.csv: A report-level dataset containing some manually checked information on immigration lobbying expenditures (1999–2007). The dataset is used in "prep-lobbyview-data.R" and contains 8 variables and 86 observations.
  - report_id: Report ID
  - client_name: Client name
  - client_legal_name: Client legal name

  - year: Report year
  - period: Report period
  - date_filed: Report filing date
  - issue: General issue area code
  - rep_amount: Total lobbying expenses in the report

- lv-check-venues-1999-2017.csv: A report-level dataset containing some manually checked information on immigration lobbying target venues (1999–2007). The dataset is used in "prep-lobbyview-data.R" and contains 8 variables and 352 observations.
  - report_id: Report ID
  - client_name: Client name
  - client_legal_name: Client legal name
  - year: Report year
  - period: Report period
  - date_filed: Report filing date
  - issue: General issue area code
  - venue: Target lobby venue(s)

- lv-other-reports-17-raw.RData: A report-level raw dataset in 2017 focusing on other select lobbying issue codes. The data were acquired directly from LobbyView's principal investigator in January 2020 and are used in "prep-lobbyview-data.R". The dataset contains 9 variables and 66,818 observations.
  - report_id: Report ID
  - client_name: Client name
  - bvdid: Client BvD ID
  - client_gvkey: Client GVKEY
  - report_amount: Total lobbying expenses in the report
  - reporting_year: Report year
  - TOB: if NULL = no reports, else the value = the list of TOB lobbying target venue(s)
  - BEV: if NULL = no reports, else the value = the list of BEV lobbying target venue(s)
  - CDT: if NULL = no reports, else the value = the list of CDT lobbying target venue(s)

- lv-all-report-issue-17-raw.RData: A report-issue-level raw dataset that contains all lobbying reports and issue codes in 2017. The data were downloaded from LobbyView's website in September 2020 and are used in "prep-lobbyview-data.R". The dataset contains 18 variables and 146,065 observations.
  - report_uuid: Unique report ID
  - client_uuid: Unique client ID

- registrant_uuid: Unique registrant ID
- registrant_name: Registrant name
- report_year: Report year
- report_quarter_code: Report quarter of the year
- amount: total lobbying expenses in the report
- is_no_activity: Whether there was no reported activity in the period?
- is_client_self_filer: Whether the filing is done in-house?
- is_amendment: Whether the report is an amendment?
- client_name: Client name
- gvkey: Client GVKEY
- bvdid: Client BvD ID
- primary_naics: Client primary NAICS
- issue_ordi: General issue area code
- issue_code: An integer indicating issue order number in the report
- issue_text: Specific lobbying issues described under the general issue area code
- gov_entity: Target lobby venue(s)

References

Kim, In Song (2018). "LobbyView: Firm-level Lobbying & Congressional Bills Database". *Working Paper*. Available at http://web.mit.edu/insong/www/pdf/lobbyview.pdf.