Codebook for the Replication Data for "Elections and the Regression-Discontinuity Design"

Devin Caughey & Jasjeet S. Sekhon

December 22, 2025

StAlphCd: Integer code indicating the alphabetical ordering of states.

CDNumAtL: Integer code indicating congressional district seat. If the seat corresponds to a geographic district (e.g., California's Fourth Congressional District), the seat is identified with that number (e.g., 4). Numbers 90 or above indicate at-large (state-wide) House seats. An at-large seat is coded 90 if it is the state's only House seat (e.g., Delaware's lone seat); 91–99 if it is one of several at-large seats and there are no district-based seats (e.g., Alabama in 1962); 990 if it is the only at-large seat but there are other district-based seats (e.g., Florida in 1932); and 991 if it is one of several at-large seats and there are also other district-based seats (e.g., Texas in 1932). States with multiple at-large seats didn't typically give them names or numbers, so the numbering of seats in the 91–99 range is essentially arbitrary.

YearElec: Year the general election was held.

Ob: Unique observation number.

Use: Dummy variable indicating whether the observation is included in the RD incumbency advantage sample. An observation was included in this sample if and only if (a) the district existed in the same form in both the preceding and subsequent election; (b) it was not an at-large seat in a state with more than one such seat; (c) the previous and subsequent elections are included the dataset; and (d) the value of the forcing variable (Democratic margin) is not missing for the observation. Corrections to the data aside, this sample is essentially identical to the sample used in Lee (2008).

CngrsNum: Congress number corresponding to the election year.

StICPSR: ICPSR code for state.

StPOAbrv: US Post Office abbreviation for state.

DifDPct: The Democratic margin of victory or defeat in the election. Calculated as the difference between the percentage of all votes that were cast for the leading Democrat in the race and the percentage cast for the leading non-Democrat. Races in which no Democrat ran or in which the top two vote-getters were both Democrats are coded as missing.

DemPct: Percentage of all votes that were cast for the leading Democrat in the race.

DemWin: Dummy variable indicating whether a Democrat won the election.

DWinPrv: Dummy variable indicating whether a Democrat won the previous election in the same district. Coded as missing if the seat was redistricted between the previous election and the current one.

DWinNxt: Dummy variable indicating whether a Democrat won the next election in the same district. Coded as missing if the seat was redistricted between the current election and the next one.

DifPVDec: Difference between the Democratic and Republican percentages of the presidential vote in the district, averaged over all elections in that decade. Data on district-level presidential vote were provided by David W. Brady.

OpenSeat: Dummy variable indicating whether the race was an open-seat election (i.e., whether the current occupant of the seat was not running for reelection). Derived from candidate data provided by Gary Jacobson.

IncStatus: Party and retirement status of the incumbent. Derived from candidate data provided by Gary Jacobson.

PrvTrmsD: Number of previous terms that the Democratic candidate had served in the U.S. House at the time of the election. Calculated by matching the candidate to Poole and Rosenthal's NOMINATE dataset.

PrvTrmsO: Number of previous terms that the main non-Democratic candidate (typically a Republican) had served in the U.S. House. Calculated by matching the candidate to Poole and Rosenthal's NOMINATE dataset (www.voteview.com).

`VoteTotW`: Total number of votes cast for the winner of the election.

`VoteTotL`: Total number of votes cast for the runner-up in the election.

`VoteCast`: Total number of votes cast for all candidates in the election.

`SoSDem`: Dummy variable indicating whether, at the time of the election, the state's secretary of state was a Democrat.

`PrvElcOb`: Observation number (`Ob`) corresponding to the previous election in this district. Coded as missing if the seat was redistricted between the previous election and the current one.

`NxtElcOb`: Observation number (`Ob`) corresponding to the next election in this district. Coded as missing if the seat was redistricted between the current election and the next one.

`DemInc`: Dummy variable indicating whether there was a Democratic incumbent running in the race. Derived from candidate data provided by Gary Jacobson.

`DemOpen`: Dummy variable indicating whether the current incumbent was a Democrat who was not running for reelection. Derived from candidate data provided by Gary Jacobson.

`NonDInc`: Dummy variable indicating whether there was a non-Democratic incumbent running in the race. Derived from candidate data provided by Gary Jacobson.

`NonDOpen`: Dummy variable indicating whether the current incumbent was a non-Democrat who was not running for reelection. Derived from candidate data provided by Gary Jacobson.

`DExpAdv`: Dummy variable indicating whether the Democratic candidate had an experience advantage over the main non-Democratic candidate. Coded as 1 only if the Democrat (a) was the incumbent representative and her opponent was not a former member of Congress or (b) had held another elected office but her opponent had not.

`RExpAdv`: Dummy variable indicating whether the Republican candidate had an experience advantage over the main Democratic candidate. Coded as 1 only if the Republican (a) was the incumbent representative and her opponent was not a former member of Congress or (b) had held another elected office but her opponent had not.

ForgnPct: Percent foreign-born in the district. Derived from U.S. Census data by Scott Adler.

GovWkPct: Percent of the district population that worked for the government. Derived from U.S. Census data by Scott Adler.

BlackPct: Percent black in the district. Derived from U.S. Census data by Scott Adler.

UrbanPct: Percent urban in the district. Derived from U.S. Census data by Scott Adler.

VtTotPct: Total votes cast in the election, as a percent of district population.

DPctNxt: Democratic percentage of the vote in the next election. Coded as missing if the seat was redistricted between the current election and the next one.

DifDPNxt: Democratic margin of victory of defeat in the next election. Coded as missing if the seat was redistricted between the current election and the next one.

DPctPrv: Democratic percentage of the vote in the previous election. Coded as missing if the seat was redistricted between the previous election and the current one.

DifDPPrv: Democratic margin of victory of defeat in the previous election. Coded as missing if the seat was redistricted between the last election and the current one.

ElcSwing: Difference between the average Democratic margin across all House races in the current election year and the average Democratic margin in the previous election (i.e., the partisan swing between this election and the last).

GovDem: Dummy variable indicating whether, at the time of the election, the state's governor was a Democrat.

IncDWNOM1: The incumbent representative's DW-NOMINATE score (obtained from Poole and Rosenthal's www.voteview.com).

DSpndPct: Campaign expenditures of the leading Democratic candidate, as a percentage of total expenditures of the leading Democrat and leading Republican. Data provided by Gary Jacobson and available since 1972. May possibly include some post-election spending.

DDonaPct: Pre-election campaign donations to the leading Democratic candidate, as a percentage of all donations received by the leading Democrat and leading Republican. Includes all direct itemized donations by individuals and committees ($\geq$\$200). Specifically, the data include transactions coded by the Federal Election Commission as 15 ("contribution" by individual), 15E ("earmarked contribution" by committee), 24K ("contribution made to non-affiliated" by committee), and 24Z ("in-kind contribution made to registered filer" by committee). For more information, see `http://www.fec.gov/finance/disclosure/metadata/DataDictionaryTransactionTypeCodes.shtml`. If a runoff was held, which was very rarely, these donations are included as well. The data are available only since 1980. Based on data provided by Jonathan Wand, who derived it from FEC donation data.

CQRating3: *Congressional Quarterly*'s pre-election prediction for the race (`-1` = Republican victory, `0` = tossup, `1` = Democratic victory). Obtained from the corresponding October issues of *CQ*.