ONLINE APPENDIX

Do Earmarks Target Low-Income and Minority Communities? Evidence from US Drinking Water

By David A. Keiser, Bhashkar Mazumder, David Molitor, Joseph S. Shapiro, and Brant J. Walker

I. Data

A. Data Sources

We use records of Safe Drinking Water Act (SDWA) loans, which we obtained from federal Freedom of Information Act requests (Office of Ground Water and Drinking Water 2019, 2023). Keiser et al. (2023) discuss the data through calendar year 2019, which this paper extends through 2022.

We also use records of earmarks, obtained from public sources (Taxpayers for Common Sense 2012; Orey, Wuerfmannsdobler and Thorning 2022). The earmark data cover fiscal years 2008 through 2010 and 2022, and they list the fiscal year of each earmark but no other date information.[1] We analyze fiscal years 2009, 2010, and 2022, which have both earmark and loan data. Our loan data begin in January 2009, so we exclude the first three months of fiscal year 2009.

From the 2010 Census, we gather information on the share of the population who is Black and Hispanic for each census block. Due to data limitations, we use the 2006–2010 average of the American Communities Survey for the share of the population below the poverty line for each census block group. Demographic information is accessed via the National Historical Geographic Information System (Manson et al. 2023). Blocks are the smallest geographic unit that the US Census Bureau identifies.

We gather population served for the roughly 150,000 active public water systems from the Environmental Protection Agency's (EPA) Safe Drinking Water Information System (SDWIS) (Envirofacts Data Service API 2020). About 50,000 of these are community water systems, which serve permanent households year round.

We use electronic maps (shapefiles) describing the areas each public water system serves, from SimpleLab and EPIC (2022). SimpleLab and EPIC (2022) work with state governments to document these maps. When creating water system boundaries, these data describe three "Tiers" of data quality. Tier 1 systems use water system boundaries from electronic maps (i.e., shapefiles). Tier 2 systems use an algorithm to identify a one-to-one match with electronic TIGER/Line maps from the Census Bureau that define a town or city boundary. Tier 3 systems use the best available system centroid and draw a circle around it using a statistical model to estimate the radius. Our main results only analyze Tiers 1 and 2 given their more accurate maps, though a sensitivity analysis adds back Tier 3.

B. Demographic Links

Our analysis sample reflects exclusions based on topic or missing key variables. Our data report on 42,354 total congressional earmarks in all domains (transportation, energy, water, etc.) in fiscal years 2009, 2010, and 2022. To determine which earmarks fund drinking water, we limit the sample to project descriptions containing the following words related to drinking water projects. These keywords were determined by manually reviewing thousands of project descriptions, especially those mentioning water (many of these fund wastewater investments under the Clean Water Act, not drinking water improvements under the SDWA). The selected keywords are "drinking," "well," "public water system," "supply," "purification," "water use," "stor-

---

[1] The Congressional fiscal year runs from October of the previous year to the September of the focal year.

age," "water tank," "storage tank," "water intake," or "water and wastewater." This process resulted in 283 projects. We then manually review these projects, using the project description and internet searches of any projects where the fund description does not clearly identify whether it is a drinking water project. This leaves 205 congressional earmarks for drinking water.

We identify 2,097 SDWA loans in fiscal years 2009, 2010, and 2022. Our analysis sample excludes 97 loans without a public water system identifying code and 11 loans targeted to territories outside the 50 states, giving a sample of 1,989.

Joining Loans and Earmarks to Demographics

Because the earmarks data have string descriptions of the recipient but not a water system identification code, we match each earmark to the water system or local government receiving it through the following steps. First, we identify the name of the state and local government receiving the earmark from the project description variable in the earmarks data. Second, we search water system names in SDWIS within the recipient state for the recipient government. If a water system exists with the local government name, we match the earmark to the system. For example, we would match an earmark from Arizona that has the description "new drinking water project for the city of Glendale" to a public water system with the name "glendale city of" in Arizona. Third, if multiple or no water systems have the targeted city name, we match the earmark to a corresponding census place, county subdivision, or county. In the previous example, if SDWIS has two public water systems with the names "glendale city of, west" and "glendale city of, east," we match the earmark to the census place for Glendale, AZ.

Finally, we double-check all public water system matches using the following steps: (i) separately identify the recipient community, using internet searches for the specific drinking water project described in the project description; (ii) overlay shapefile maps from SimpleLab and EPIC with open street maps; and (iii) identify the shapefile for the matched public water system from our hand-matching algorithm, and ensure the community identified in step (i) is within its borders. Out of 205 earmarks satisfying earlier sample exclusions, we match 163 to a public water system and 40 to a local government, and further exclude 3 targeted to the entire state of Alaska and Puerto Rico. The main text refers to water systems, local governments, or census areas receiving earmarks or loans collectively as communities or water systems.

We then join water systems to demographics, which we collect from Manson et al. (2023). We extract the latitude and longitude of the centroid of each census block from Manson et al. (2023). Using the water system service territory maps, we spatially join each of these census block centroids with the water system(s) serving that census block. If a block's centroid is located within the boundary of multiple water systems, we include it in each system for calculating demographics. We then assign water systems the population-weighted demographics of all census blocks they serve. Using the census block data, we calculate the share of the population in each system or local government who is Black, Hispanic, or poor. If an earmark links to a local government but not a water system, we use census demographics for the census place or county subdivision, also from Manson et al. (2023). Our final sample with demographic information includes about 39,000 public water systems; nearly all are community water systems. We match 157 earmarks to demographics, 141 of which have Tier 1 or 2 shapefile boundaries. We match 1,855 SDWA loans to demographics, 1,579 of which have Tier 1 or 2 shapefile boundaries.

II. Additional Background

The 1996 SDWA amendments created the State Revolving Fund that we study. Regularly since 1997, Congress has appropriated annual funding for it. In most years, Congress dedicates about $1 billion to states as capitalization grants. In fiscal

year 2009, the American Recovery Act increased this number temporarily to nearly $3 billion. Starting in fiscal year 2022, the 2021 Infrastructure Investment and Jobs Act increased appropriations to over $6 billion annually through fiscal year 2026 (Tiemann 2023).

Until the ban on traditional earmarks in fiscal year 2011, Congress appropriated earmarks for drinking water programs separately from the state capitalization grants, though the EPA distributed them through the same account. In fiscal year 2022, Congress changed this practice and allocated a portion of the annual capitalization grants to earmarks (Tiemann 2023).

The EPA and states can set aside a portion of annual appropriations for specific projects. Congress usually directs the EPA to allocate a small percentage of appropriations for tribal nations, US territories, and special programs such as operator training and unregulated contaminants (Congressional Research Service 2018). The SDWA allows states to reserve about 30% of funds for needs including funding for disadvantaged communities, technical assistance and training, and source water protection programs (Congressional Research Service 2018). It requires individual states to provide a 20% match on the remaining capitalization funds.[2] The SDWA also requires that states allocate at least 15% of annual funds for water systems serving 10,000 or fewer individuals (Congressional Research Service 2018).

The main text mentions over $50 billion in loan spending. This statistic reflects EPA calculations of total program-related expenditures from 1997 to 2021. During this period, the EPA provided $25 billion in grants to states, which supported $53 billion in total funds to water systems (EPA 2023a). The main text also notes that the 2021 Infrastructure Investment and Jobs Act allocates $40 billion to drinking water infrastructure; this number comes from EPA (2023b). The agency notes the 2021 Act allocates $11.7 billion in Drinking Water State Revolving Fund general spending,

$15 billion in similar loans for lead service line replacement, and $4 billion in loans and $5 billion in grants to address emerging contaminants. Another $12.7 billion is allocated to address wastewater needs (EPA 2023b).

III. Additional Summary Statistics and Additional Results

Appendix Table A1 shows summary statistics on loans, earmarks, and demographics. Four percent of water systems in our data receive one or more SDWA loans in the fiscal years we study, and half a percent receive earmarks. In the mean community, 7% of people are Black, 10% are Hispanic, and 14% have income below the poverty line. The mean system receiving a loan or earmark serves 60,000 to 70,000 people, though the mean system overall serves about 8,000 people.

The map in Appendix Figure A1 shows which counties have water systems that receive loans, earmarks, or both. All states have systems that receive SDWA loans, and most have systems that receive drinking water earmarks. Earmarks and loans cover both rural and urban areas, though counties located within major metro areas are more likely to receive both.

Appendix Table A2 shows sensitivity analyses that examine the relationship between demographics, earmarks, and SDWA loans, as reported in Table 1. Many patterns echo the main results, but we comment on a few of the differences. Democratic earmarks are more likely to target Black and Hispanic communities, but Republican earmarks are more likely to target poor communities. After controlling for population size, we observe no statistical difference in the targeting of Black communities by earmarks and loans. Controlling for state fixed effects, loans target Hispanic communities at roughly the same rate as earmarks. Finally, including lower-quality shapefiles that match public water systems to census demographics do not affect our main findings.

[2]Subpart L of Title 40 of the Code of Federal Regulations (40 CFR 35.3560 2023) defines the federal share as net of set-asides.

APPENDIX REFERENCES

40 CFR 35.3560. 2023. "Title 40, Code of Federal Regulations, part 35, section 3560."

Congressional Research Service. 2018. "Drinking Water State Revolving Fund (DWSRF): Overview, Issues, and Legislation." R45304.

Envirofacts Data Service API. 2020. "Safe Drinking Water Information System - Water Systems." US EPA. `https://www.epa.gov/enviro/envirofacts-data-service-api` (accessed May 13th, 2020).

EPA. 2023*a*. "Drinking Water State Revolving Fund Infographic." `https://www.epa.gov/dwsrf/drinking-water-state-revolving-fund-infographic`, accessed December 10, 2023.

EPA. 2023*b*. "Water Infrastructure Investments." `https://www.epa.gov/infrastructure/water-infrastructure-investments`, accessed December 10, 2023.

Keiser, David A., David Bhashkar Mazumder, David Molitor, and Joseph S. Shapiro. 2023. "Water Works: Causes and Consequences of Safe Drinking Water in America." `https://joseph-s-shapiro.com/research/DrinkingWaterPollution.pdf`.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2023. "IPUMS National Historical Geographic Information System: Version 18.0." Minneapolis, MN: IPUMS. [2006_2010_ACS5a; 2010_SF1a; 2010 TIGER/Line+]. http://doi.org/10.18128/D050.V18.0.

Office of Ground Water and Drinking Water. 2019. "Freedom of Information Request." United States Environmental Protection Agency. (Received August 7th, 2019).

Office of Ground Water and Drinking Water. 2023. "Freedom of Information Request." United States Environmental Protection Agency. (Received October 31st, 2023).

Orey, Rachel, Franz Wuerfmannsdobler, and Michael Thorning. 2022. "Congressionally Directed Spending FY2022 Dataset." Bipartisan Policy Center. `https://bipartisanpolicy.org/blog/congressionally-directed-spending-fy2022-dataset/` (accessed October 2nd, 2023).

SimpleLab, and EPIC. 2022. "U.S. Community Water Systems Service Boundaries, v3.0.0." HydroShare. `http://www.hydroshare.org/resource/9ebc0a0b43b843b9835830ffffdd971e` (accessed February 6th, 2023).

Taxpayers for Common Sense. 2012. "All Earmarks for FY2008-FY2010." `https://www.taxpayer.net/budget-appropriations-tax/earmark-data/` (accessed October 13th, 2023).

Tiemann, Mary. 2023. "The Role of Earmarks in CWSRF and DWSRF Appropriations in the 117th Congress." Congressional Research Service R47633.

Figure A1: Map of Federal Funds for Drinking Water Projects


Received Earmarks
Received SDWA loans
Received both
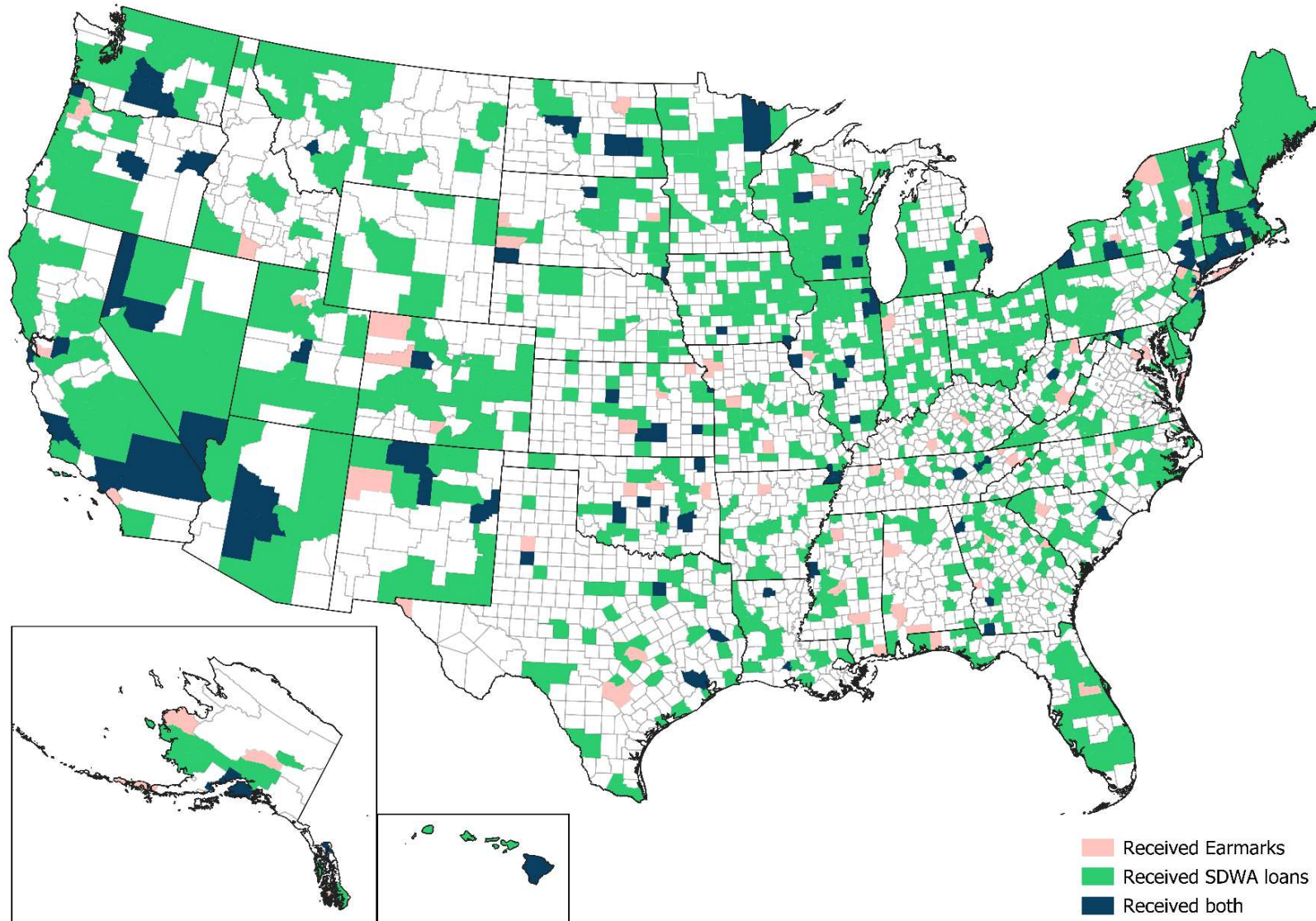
Notes: This map indicates whether each county had a system that received earmarks, Safe Drinking Water Act loans, or both, in fiscal years 2009, 2010, and 2022.

Figure A2: Targeting of Earmarks and Loans, by Community Demographics, Controlling for Log Population

Panel A: Community share Black
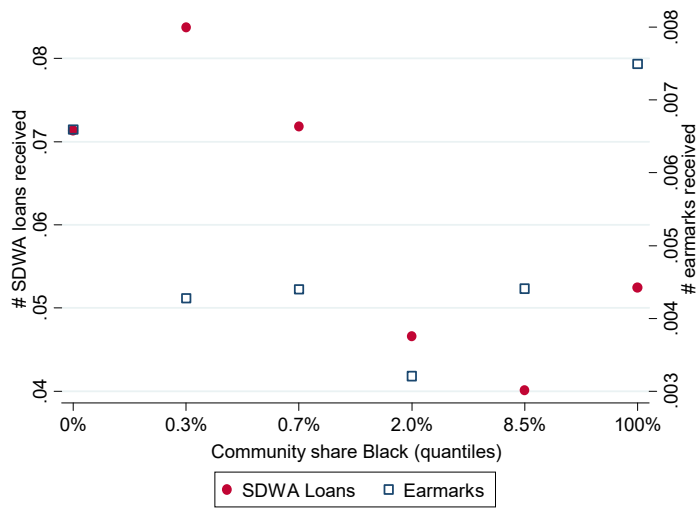
Panel B: Community share Hispanic
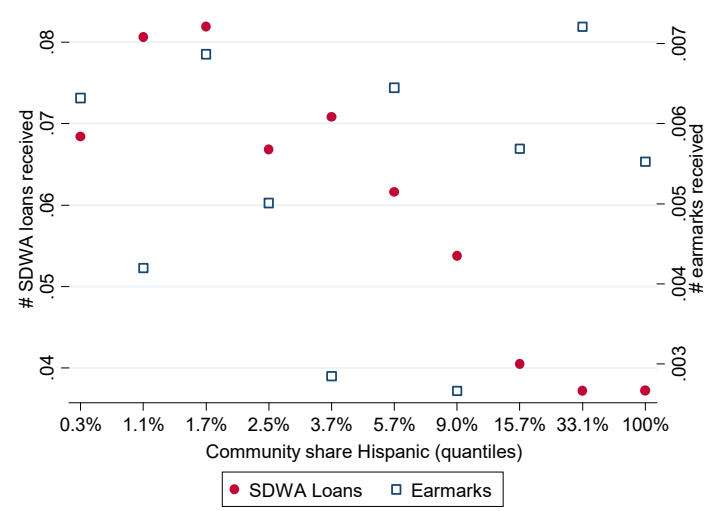
Panel C. Community poverty rate
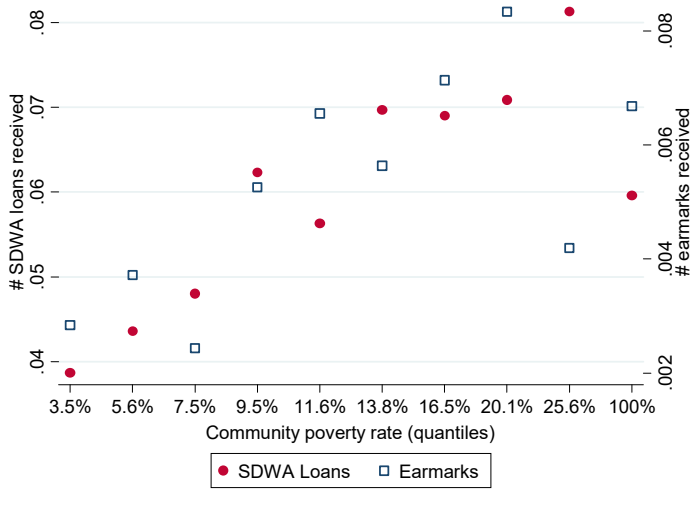
Notes: An observation underlying the graphs represents a water system. SDWA is the Safe Drinking Water Act. Each panel divides water systems into deciles based on the relevant community demographic. Each point in a graph shows the number of loans or earmarks that water systems receive, divided by the total number of water systems in the decile. Panel A shows six bins due to the large number of communities with zero share Black. In each graph, the horizontal axis labels show the maximum value in each bin (Panels A-C) or the decile number (Panel D). Demographics are from the 2010 Census (Manson et al. 2023).

Table A1: Summary Statistics

A. Safe Drinking Water Act Loans	
Total number of loans	1,989
Mean population served	69,123
Share of systems with a loan	0.04
B. Congressional Earmarks	
Total number of earmarks	203
Mean population served	62,301
Share of systems with an earmark	0.004
C. Census Demographics	
Systems with demographics	39,182
Share Black (%)	7.17%
	(0.16)
Share Hispanic (%)	9.67%
	(0.17)
Share in poverty (%)	13.66%
	(0.10)
Mean population served	7,979
	(69,005)
Mean log population	6.75
	(1.92)

Notes: The table shows mean values. Standard deviations are in parentheses. Each observation represents one water system. The number of loans and earmarks, and the share of systems receiving funds, include all funds awarded in fiscal years 2009, 2010, and 2022. Demographics are from the 2010 Population Census (Manson et al. 2023). Data on the population served by each water system is from SDWIS (Envirofacts Data Service API 2020).

Table A2: Targeting of Drinking Water Earmarks and Loans, by Community Demographics: Sensitivity Analyses

	Earmarks (1)	SDWA Loans (2)	p-val: (1)=(2) (3)
Panel A. Share Black			
1. Baseline (main result)	17.545	78.439	[0.000]
	(5.527)	(15.497)	
N	26,529	26,529	
2. Any loan or earmark (x1,000)	16.519	59.217	[0.001]
	(4.886)	(11.328)	
N	26,529	26,529	
3. Any loan or earmark - logistic regression (x1,000)	10.396	46.953	[0.067]
	(2.049)	(7.252)	
N	26,529	26,529	
4. Earmarks by Democratic lawmakers	12.619	—	—
	(4.932)	—	
N	26,529	—	
5. Earmarks by Republican lawmakers	8.704	—	—
	(3.351)	—	
N	26,529	—	
6. Add state fixed effects	22.301	138.798	[0.000]
	(6.754)	(21.146)	
N	26,529	26,529	
7. Analyze earmarks and loans in same regression	17.100	77.414	[0.000]
	(5.515)	(15.508)	
N	26,529	26,529	
8. Weight by population served	-37.629	501.169	[0.107]
	(106.016)	(306.070)	
N	26,471	26,471	
9. Control for log population served	2.866	5.377	[0.875]
	(5.577)	(15.058)	
N	26,471	26,471	
10. Control for urban/rural	14.201	59.955	[0.004]
	(5.157)	(15.201)	
N	26,491	26,491	
11. Include Tier 1, 2, and 3 shapefiles	8.503	19.651	[0.206]
	(3.060)	(8.367)	
N	39,059	39,059	

Continued on next page

Table A2: Targeting of Drinking Water Earmarks and Loans, by Community Demographics: Sensitivity Analyses

	Earmarks (1)	SDWA Loans (2)	p-val: (1)=(2) (3)
Panel B. Share Hispanic			
1. Baseline (main result)	11.142	-46.896	[0.000]
	(4.045)	(7.576)	
N	26,529	26,529	
2. Any loan or earmark (x1,000)	6.995	-40.810	[0.000]
	(2.895)	(5.741)	
N	26,529	26,529	
3. Any loan or earmark - logistic regression (x1,000)	5.537	-51.042	[0.000]
	(1.834)	(8.993)	
N	26,529	26,529	
4. Earmarks by Democratic lawmakers	12.232	—	—
	(3.737)	—	
N	26,529	—	
5. Earmarks by Republican lawmakers	4.059	—	—
	(3.022)	—	
N	26,529	—	
6. Add state fixed effects	20.434	34.660	[0.143]
	(4.792)	(8.425)	
N	26,529	26,529	
7. Analyze earmarks and loans in same regression	11.433	-47.605	[0.000]
	(4.037)	(7.590)	
N	26,529	26,529	
8. Weight by population served	580.010	87.582	[0.320]
	(237.027)	(377.249)	
N	26,471	26,471	
9. Control for log population served	7.300	-61.909	[0.000]
	(3.756)	(7.544)	
N	26,471	26,471	
10. Control for urban/rural	0.444	-71.977	[0.000]
	(2.626)	(7.921)	
N	26,491	26,491	
11. Include Tier 1, 2, and 3 shapefiles	11.487	-27.570	[0.000]
	(3.412)	(6.407)	
N	39,059	39,059	

Continued on next page

Table A2: Targeting of Drinking Water Earmarks and Loans, by Community Demographics:
Sensitivity Analyses

	Earmarks (1)	SDWA Loans (2)	p-val: (1)=(2) (3)
Panel C. Share in poverty			
1. Baseline (main result)	22.682	125.745	[0.000]
	(6.080)	(17.984)	
N	26,508	26,508	
2. Any loan or earmark (x1,000)	17.898	95.894	[0.000]
	(5.024)	(13.645)	
N	26,508	26,508	
3. Any loan or earmark - logistic regression (x1,000)	14.781	84.973	[0.359]
	(3.384)	(10.733)	
N	26,508	26,508	
4. Earmarks by Democratic lawmakers	9.239	—	—
	(5.067)	—	
N	26,508	—	
5. Earmarks by Republican lawmakers	20.916	—	—
	(4.515)	—	
N	26,508	—	
6. Add state fixed effects	20.047	163.346	[0.000]
	(6.337)	(20.077)	
N	26,508	26,508	
7. Analyze earmarks and loans in same regression	21.965	124.417	[0.000]
	(6.040)	(17.975)	
N	26,508	26,508	
8. Weight by population served	80.447	2082.208	[0.001]
	(160.536)	(585.005)	
N	26,450	26,450	
9. Control for log population served	13.177	80.312	[0.000]
	(5.885)	(17.235)	
N	26,450	26,450	
10. Control for urban/rural	17.827	134.822	[0.000]
	(5.186)	(18.172)	
N	26,470	26,470	
11. Include Tier 1, 2, and 3 shapefiles	12.414	62.366	[0.000]
	(3.591)	(11.339)	
N	39,035	39,035	

Continued on next page

Table A2: Targeting of Drinking Water Earmarks and Loans, by Community Demographics: Sensitivity Analyses

Note: Each table entry shows a separate regression. The dependent variable measures the number of loans or earmarks received per thousand systems, unless otherwise stated. The independent variable measures one community characteristic (e.g., share Black). Each observation represents one public water system. Estimates include Census block demographics matched with Tier 1 and 2 shapefiles from SimpleLab and EPIC (2022), unless otherwise stated. Heteroskedastic-robust standard errors are in parentheses. Column (3) shows the p-value from a Wald test of the null hypothesis that columns (1) and (2) have equal coefficients. Logistic regressions display the marginal effect of the independent variable and the Wald test p-value comparing the coefficients.