Overview

- **Research Questions**
  - How has water pollution changed since before 1972?
  - Did Clean Water Act cause changes?
  - How do residents value changes?
- **1972 US Clean Water Act**
  - Costly, controversial, understudied
  - *SWANCC*, *Rapanos*, Clean Water Rule, Apr 2017 Supreme Court
  - Benefits $<$ costs? (Lyon and Farrow 1995, Freeman 2000, EPA 2000)
- **Approach**
  - Analyze $680 billion in Clean Water Act expenditures
  - Many datasets new to economic research

Overview: Results

1. **Pollution has declined substantially**
2. **Clean Water Act grants caused some of this**
   - Annual cost to make a river-mile fishable: \$1.5 million
   - Complete pass-through of grants to local spending (no crowd out)
   - Limited heterogeneity in cost-effectiveness
3. **Change in housing values $<$ costs**
   - Mean ratio of measured housing benefits to costs: 0.25
   - Larger ratios in areas where fishing is popular
   - Interpretation: Lower bound? Incomplete info, general eqm, nonuse values

Motivation: Water Pollution

- **A textbook externality (Stigler 1952, 1966)**

Cuyahoga River, Ohio, 1952

Motivation: Water Pollution

- **A textbook externality (Stigler 1952, 1966)**

Cuyahoga River, Ohio, 1952

  - Ruckelshaus (first EPA head): "Even if our rivers are not fishable or swimmable, at least they're not flammable."

Existing Research and Contributions

What is new here?

- Water Pollution Trends (Smith et al. 1983; USEPA 2000; Smith & Wolloh 2012)
  - New here: national pollution changes since pre-1972
- Effects of the Clean Water Act (Earnhart 2004a,b; Robotyagov et al. 2014)
  - New here: effects on ambient pollution
- Willingness-to-pay for water quality (Smith & Desvouges 1986; Hausman et al. 1995; Keiser 2015; Mendelsohn et al. 1992; Leggett & Bockstael 2000; Walsh et al. 2011; Carson & Mitchell 1993; Kling et al. 2012; Hausman 2012)
  - New here: effects of Clean Water Act on home values
- Subsidies for pollution control equipment
  - Common policy globally, some existing theory (Kohn 1992, Aidt 1998, Fredriksson 1998), first empirical analysis

Overview

- **Background: water pollution and regulation**
- Data
- Pollution trends
- Clean Water Act and pollution
- Clean Water Act and home prices
- Conclusions

Clean Water Act Background

- **Goals:**
  - All waters safe for fishing and swimming by 1983
  - Zero water pollution discharge by 1985
- **Activity 1: Grants for municipal wastewater treatment plants**
  - In mid 1970s, largest public works program in U.S.
  - Focus of analysis
- **Activity 2: Industrial Regulation**
  - National Permit Discharge Elimination System (NPDES)
  - Controls; not the focus of analysis

Clean Water Act Background

- **Wastewater treatment plants**
  - About 15,000 plants, unchanged since 1970s
  - Goal: process wastewater before discharge to river/lake/ocean

Los Angeles Hyperion Treatment Plant

Clean Water Act Background

- **Research Design: upstream versus downstream**
- **Indirect tests of validity**
  - Event study
  - Compare across pollutants
  - Add controls: nonattainment, NPDES, population
  - DDD and DD estimators

Grant Distribution Rules

Water Pollution Background

- **Dissolved oxygen deficit**
- **Fishable**
- **Other pollutants**:
  - Swimmable
  - Biochemical oxygen demand (BOD)
  - Fecal coliforms
  - Total suspended solids (TSS)

Perceptibility

Overview

- Background: water pollution and regulation
- **Data**
- Pollution trends
- Clean Water Act and pollution
- Clean Water Act and home prices
- Conclusions

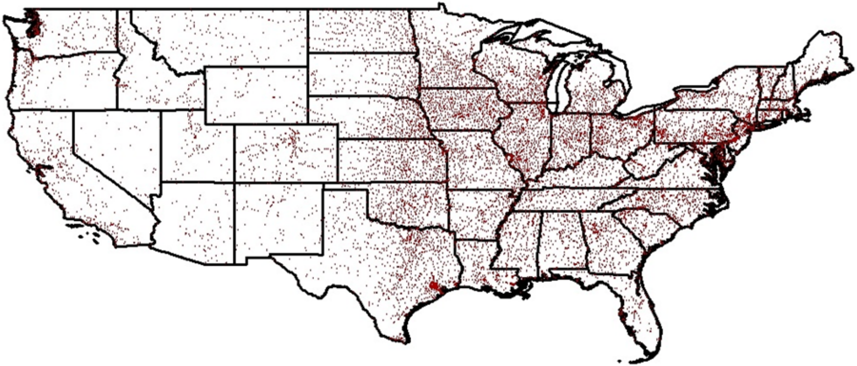
Data

- **Treatment plants; grants; river network; water pollution; population/housing census; industrial sources; recreational distances; municipal spending; other environmental data**
- **Data hard to get**
  - Freedom of Information Act requests
  - C++ with hydrology software developer
  - Interviews with EPA, USGS, private engineers

Data: Municipal Pollution Sources

- **Clean Watershed Needs Survey 1976-1996 Panel**

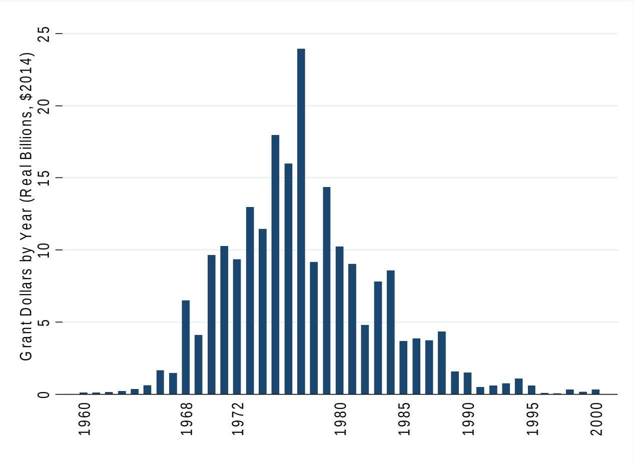
  Trends over time



Data: Clean Water Act Grants

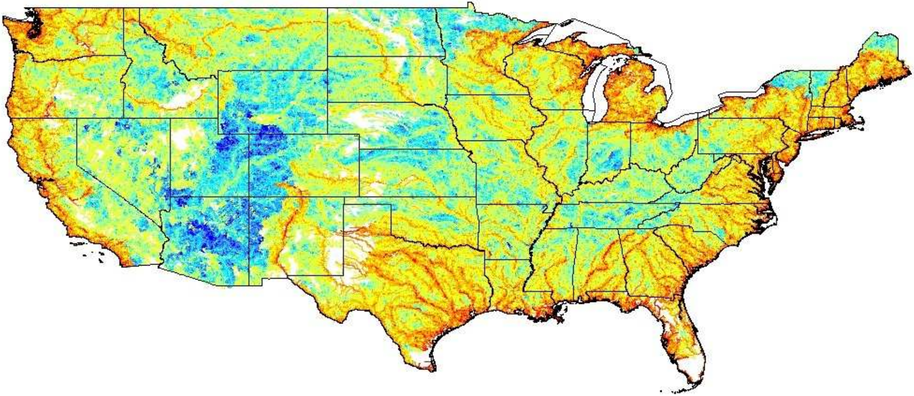
- **Grants Information and Control System (archived version)**
  - 35,000 grants

Figure: Clean Water Act Grant Dollars ($2014)



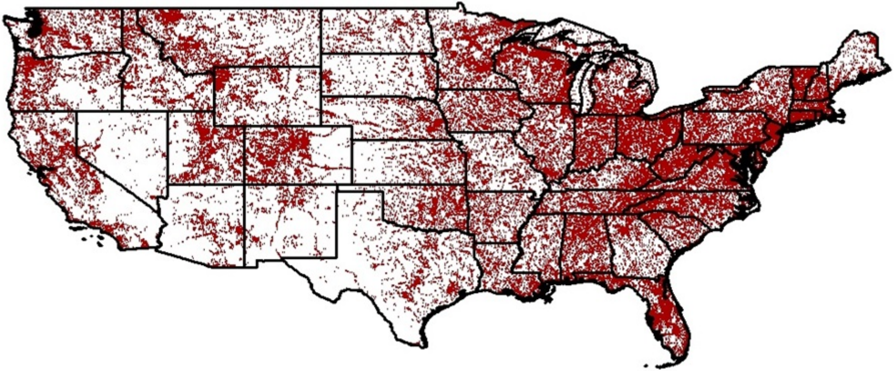
Data: Spatial Rivers and Lakes

- **National Hydrography Dataset Plus, v2.1**
  - 70 million river nodes, flow network



Data: Ambient Pollution

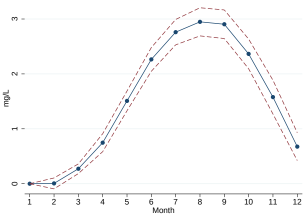
- **Combine three repositories**
  - Storet Legacy; Modernized Storet; National Water Information System
  - 50 million observations, 170,000 monitors, years 1962-2001



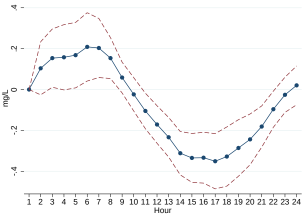
- Sample selection and cleaning. Limit to routine ambient monitoring; water samples; rivers/streams and lakes; surface water; comparable parameter codes; non-missing date, latitude, longitude; winsorize at 99th percentile.

Data: Ambient Pollution

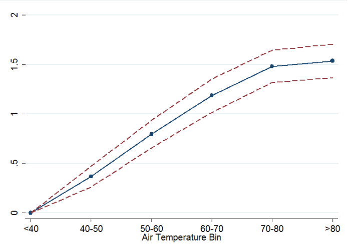
(a) By Month



(b) By Hour



(c) By Air Temperature Bin

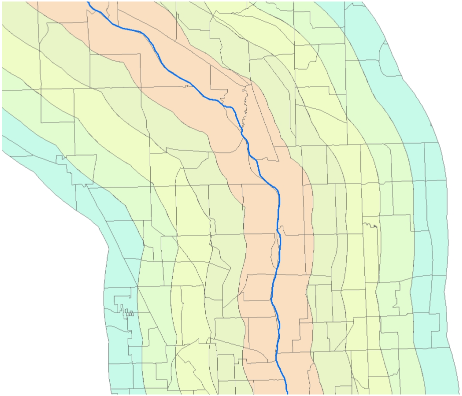


Data: Industrial Water Pollution

- **Survey of Water Use in Manufacturing (SWUM)**
  - Confidential, at Census Research Data Center (RDC)
  - Number of manufacturing plants in 1972 with large water use
- **Permit Compliance System (PCS)**
  - Cumulative permitted plants in a county, by year

Data: Census of Housing and Population

- **Geolytics Neighborhood Change Database 1970-2000**
  - Mean home values and rents, by tract



Mississippi River in Northern Minneapolis

Recreational Distances

- **Nationwide Personal Travel Survey**
  - 95th percentile of recreational trip distance: 34 miles
- **Mean ratio of road to straight-line distance: 1.4**
  - NHTS (2009), Boscoe et al. (2012)
- **Final radius: 25 miles ($\approx$ 34/1.4)**

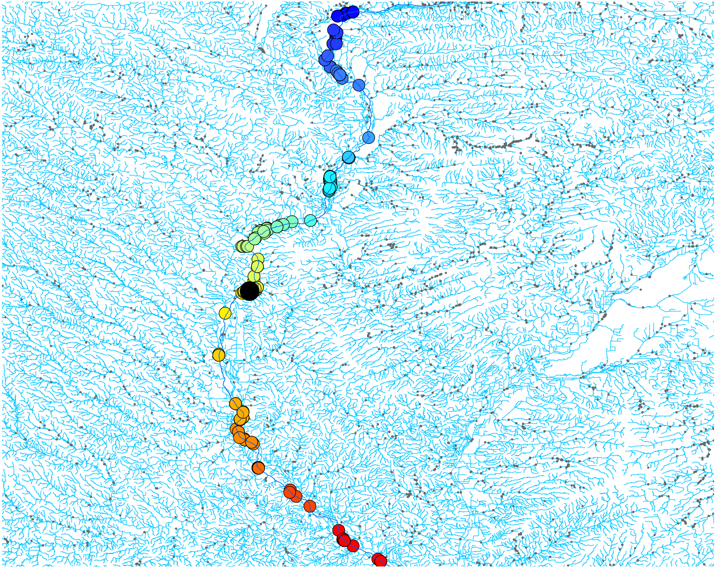
Municipal Spending

- **Census of Governments, Annual Survey of Governments**
  - Total annual investment in sewerage capital
  - 1972-2001, balanced panel of 198 cities

Data: Environmental Controls

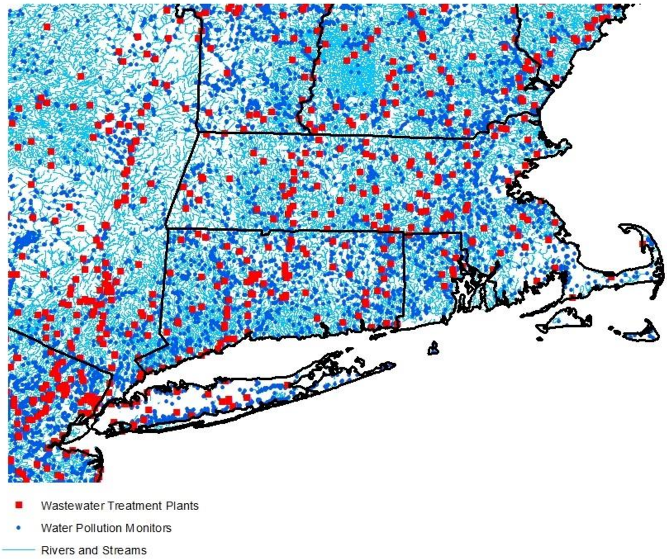
- **Temperature and Precipitation**
  - National Climate Data Center Summary of the Day files (TD-3200)
- **Air pollution regulation**
  - Nonattainment designations, by pollutant

Data: Summary Picture



Mississippi River, Northern St. Louis

Data: Summary Picture, Many Plants



Overview

- Background: water pollution and regulation
- Data
- **Pollution trends**
- Clean Water Act and pollution
- Clean Water Act and home prices
- Conclusions

Pollution Trends: Econometrics

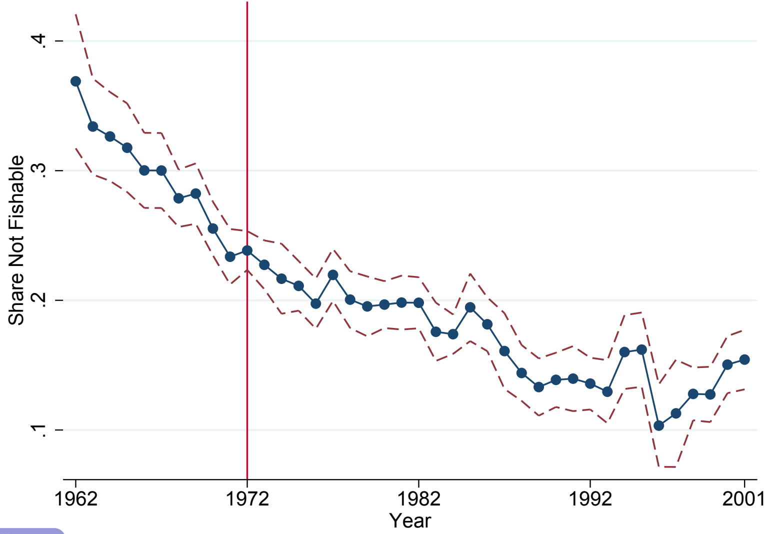
$$Q_{icy} = \alpha Y_y + X'_{icy}\beta + \delta_i + \epsilon_{icy}$$

$$Q_{icy} = \sum_{\tau=1963}^{\tau=2001} \alpha_\tau 1[Y_y = \tau] + X'_{icy}\beta + \delta_i + \epsilon_{icy}$$

- Station i, year y, date c
- X_{icy}: hour-of-day, day-of-year
- Cluster by watershed
  - Watershed $\equiv$ 8-digit Hydrologic Unit Code ("cataloging unit")
  - Land area in which all water flows to a point

Pollution Trends: Graph (Not Fishable)

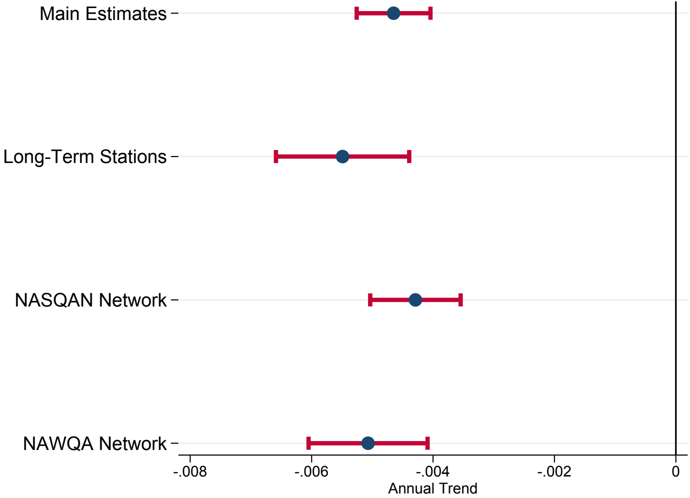
$$Q_{icy} = \Sigma_{\tau=1963}^{\tau=2012} \alpha_\tau 1[y=\tau] + X_{icy}^{'} \beta + \delta_i + \epsilon_{icy}$$



Other Pollutants

Pollution Trends: Sensitivity Analyses (Not Fishable)

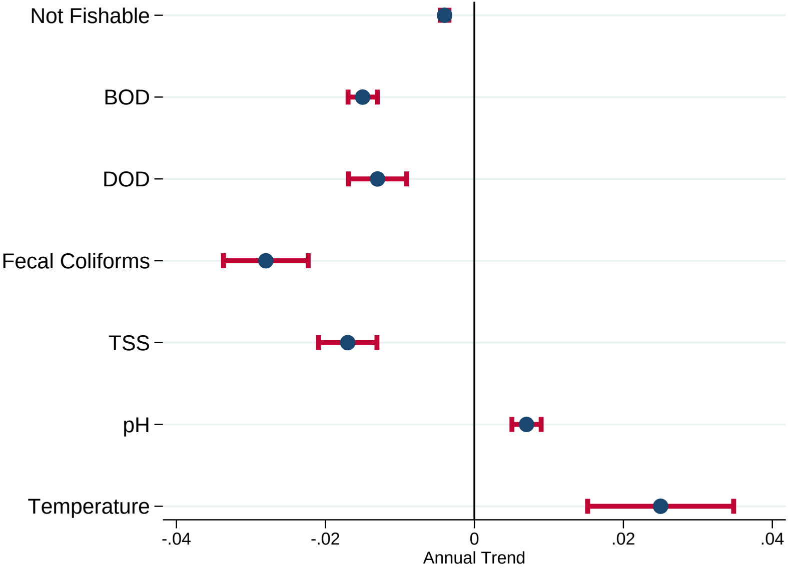
$$Q_{icy} = \alpha Y_y + X'_{icy}\beta + \delta_i + \epsilon_{icy}$$



- Additional analyses in paper: (1) Areas with home values data; (2) Drop infrequent stations; (3) Streamflow controls; (4) July-August only; (5) Exclude dam areas; (6) Recode below-detectable-limit; (7) Levels, not logs; (8) River-year means; (9) 50% fishable cutoff; (10) Water quality index; (11) Two-way cluster; (12) Lakes; (13) Population weighted

Pollution Trends: Other Pollutants

$Q_{icy} = \alpha Y_y + X_{icy}\beta + \delta_i + \epsilon_{icy}$



Overview

- Background: water pollution and regulation
- Data
- Pollution trends
- **Clean Water Act and pollution**
  - **Water quality**
  - Municipal spending
- Clean Water Act and home prices
- Conclusions

Grants and Pollution: Econometrics

$$Q_{pdy} = \gamma G_{py} d + X'_{pdy}\beta + \eta_{pd} + \eta_{py} + \eta_{dwy} + \epsilon_{pdy}$$

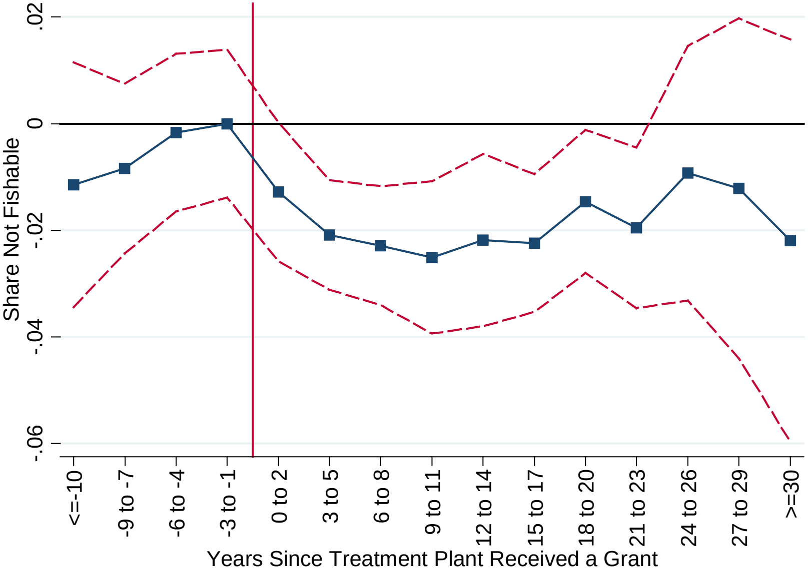
- Plant p, downstream d, year y
- G_{py}: cumulative number of grants for a plant
- η_{pd}: Plant-by-downstream fixed effects
- η_{py}: Plant-by-year fixed effects
- η_{dwy}: Downstream-by-hydrologic-region-by-year fixed effects
- Assumption:

$$E[G_{py} d \cdot \epsilon_{pdy} | \eta_{pd}, \eta_{py}, \eta_{dwy}, X'_{pdy}] = 0$$

Cross-sectional upstream v. downstream

Grants and Pollution: Event Study

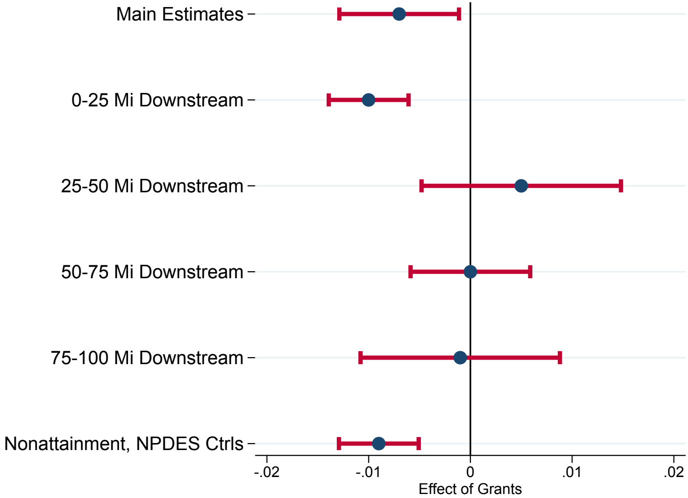
$$Q_{pdy} = \Sigma_{\tau=-10}^{\tau=25} \gamma_\tau 1[G_{p,y+\tau} = 1]1[d = 1] + X_{iy}^{'}\beta + \eta_{pd} + \eta_{py} + \eta_{dwy} + \epsilon_{pdy}$$



Other Pollutants

Grants and Pollution: Sensitivity Analyses (Not Fishable)

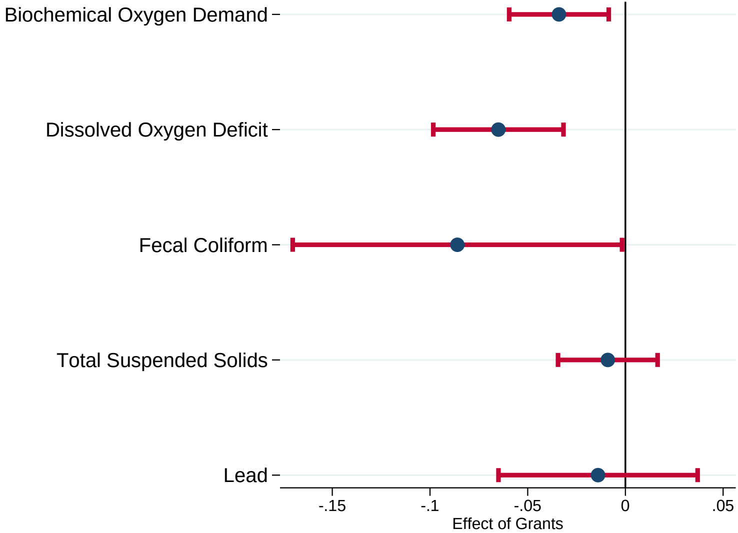
$$Q_{pdy} = \gamma G_{pdy} d + X_{pdy}^{'} \beta + \eta_{pd} + \eta_{py} + \eta_{dwy} + \epsilon_{pdy}$$



- Additional analyses in paper: (1) All 16 sensitivity analyses reported for trends; (2) Plants with monitors $\geq$ 10 mi. upstream and downstream; (3) Grants for construction; (4) Dummies for # grants; (5) Control for cumulative upstream grants; (6) Cumulative real grant dollars; (7) Diff-in-diff, downstream only; (8) Include monitors on other upstream/downstream rivers; (9) Monitor-level, station FE; (10) Monitor-level, downweight duplicates

Grants and Pollution: Other Pollutants

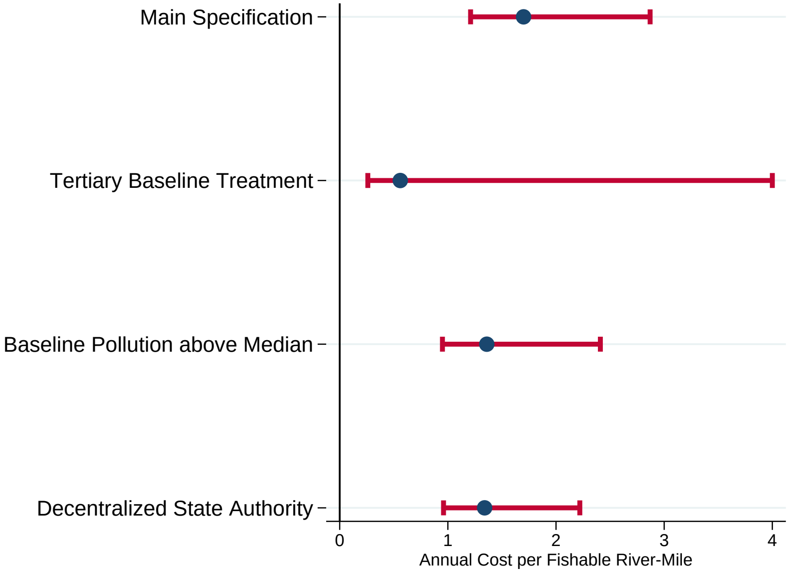
$$Q_{pdy} = \gamma G_{pdy} d + X'_{pdy}\beta + \eta_{pd} + \eta_{py} + \eta_{dwy} + \epsilon_{pdy}$$



Grants and Pollution: Cost Effectiveness

	(1)	(2)	(3)
Annual Cost to Make a River-Mile Fishable	1.82	1.40	1.51
	[1.29 , 3.08]	[1.00 , 2.37]	[1.00 , 3.07]
Total Costs	295,450	395,162	618,301
Federal Capital	87,551	117,222	184,306
Local Capital	37,167	49,773	77,118
Operation & Maintenance	170,733	228,167	356,878
River-Miles Made Fishable	5,409	9,377	16,385
Regression Sample	Yes		
All Plants		Yes	
Assumed Stream Lengths			Yes

Grants and Pollution: Cost Effectiveness Heterogeneity



Overview

- Background: water pollution and regulation
- Data
- Pollution trends
- **Clean Water Act and pollution**
  - Water quality
  - **Municipal spending**
- Clean Water Act and home prices
- Conclusions

Clean Water Act Grants and Municipal Spending

- **How did CWA grants affect wastewater treatment spending?**
  - Our measure: reported project costs
  - Crowding out? Crowding in?

- **Approach: estimate effects on municipal wastewater treatment investment**

$$E_{cy} = \beta D_{cy} + \upsilon_c + \eta_{wy} + \epsilon_{cy}$$

Clean Water Act Grants and Municipal Spending

$$E_{cy} = \beta D_{cy} + \upsilon_c + \pi_y + \eta_{wy} + \epsilon_{cy}$$

	(1)	(2)	(3)	(4)
Grant Project Costs	1.09	0.94	0.86	0.91
	(0.19)	(0.15)	(0.18)	(0.21)
City FE and Year FE	Yes	Yes	Yes	Yes
Real Costs		Yes	Yes	Yes
Basin-by-Year FE			Yes	Yes
Propensity Score Reweight			Yes	Yes

Overview

- Background: water pollution and regulation
- Data
- Pollution trends
- Clean Water Act and pollution
- **Clean Water Act and home prices**
- Conclusions

Housing: Theory

$$P_i = P(z_1, \ldots, z_J)$$

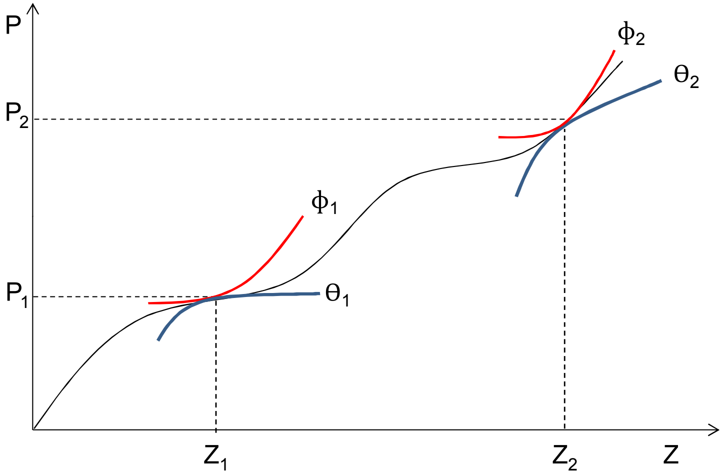
- i **House**
- P_i **Price**
- $z_1, \ldots, z_J$ **Attributes**
  - Dwelling features (e.g., square footage)
  - Local public goods (e.g., school quality)
  - Environmental amenities (e.g., clean rivers)
- $P(\cdot)$ **Hedonic price schedule**
  - Equilibrium of firms supplying housing, consumers consuming it
  - Assume competitive housing markets, each consumer rents one house

Housing: Theory

- **Marginal implicit price of attribute j : $\frac{\partial P}{\partial z_j}$**
    - Marginal: partial deriviative, holding other attributes fixed
    - Implicit: can't buy attributes separate from a house
- **Reflects tangency between consumer bid functions and firm offer functions**
- **Reveals marginal willingness to pay for amenity**

Housing: Theory

Figure: Hedonic Model



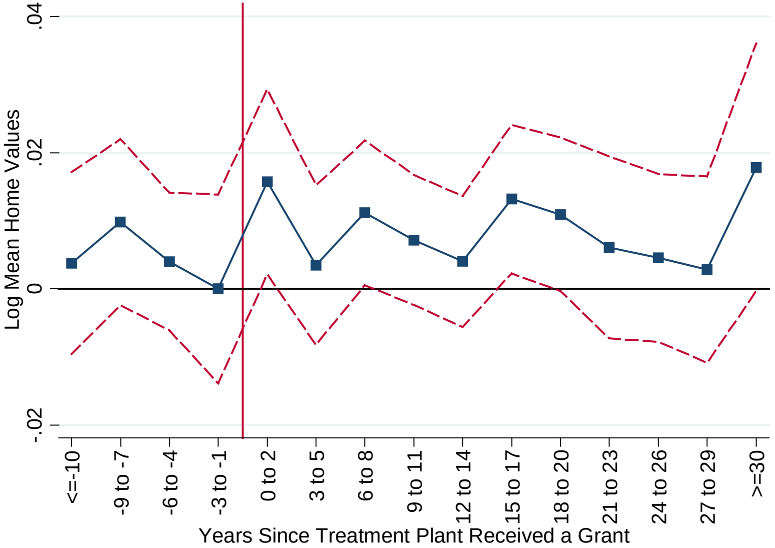
Housing: Econometrics

$$V_{pt} = \gamma G_{pt} + X'_{pt}\beta + \eta_p + \eta_{wt} + \epsilon_{pt}$$

- Plant p, year t
- G_{pt}: cumulative number of grants for a plant
- X'_{pt}: housing structure characteristics
- η_p: Plant fixed effects
- η_{wt} : Region-by-year fixed effects
- Generalized least squares. Weights reflect census sampling probabilities and housing units in an observation
- DD, not DDD

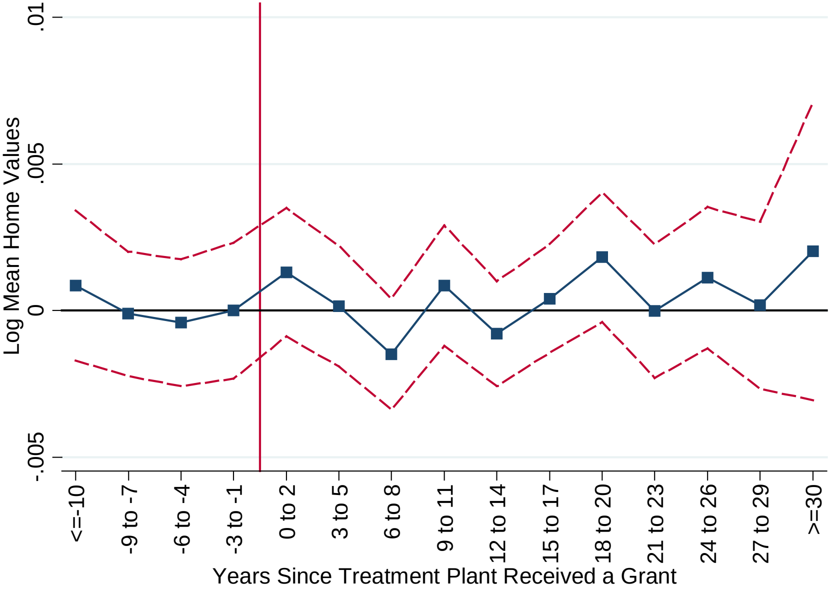
Housing: Event Study, 0.25 Mile Bandwidth

$$V_{pt} = \gamma G_{pt} + X_{pt}^{'}\beta + \eta_p + \eta_{wt} + \epsilon_{pt}$$



Housing: Event Study, 25 Mile Bandwidth

$$V_{pt} = \gamma G_{pt} + X_{pt}^{'}\beta + \eta_p + \eta_{wt} + \epsilon_{pt}$$



Housing: Tables

$$V_{pt} = \gamma G_{pt} + X_{pt}^{'}\beta + \eta_p + \eta_{wt} + \epsilon_{pt}$$

	(1)	(2)	(3)	(4)
Panel A. Log Mean Home Values				
Cumulative Grants	-0.00007	0.00090	0.002665**	0.00025
	(0.002503)	(0.001412)	(0.001286)	(0.000326)
Panel B. Log Mean Rental Values				
Cumulative Grants	0.00002	-0.00077	0.00004	-0.00011
	(0.001700)	(0.000833)	(0.000711)	(0.000158)
Plant FE, Basin-by-Year FE	Yes	Yes	Yes	Yes
Dwelling Characteristics		Yes	Yes	Yes
Baseline Covariates * Year		Yes	Yes	Yes
Max Distance Homes to River: 0.25 Miles	Yes	Yes		
Max Distance Homes to River: 1.0 Miles			Yes	
Max Distance Homes to River: 25 Miles				Yes

Housing: Tables

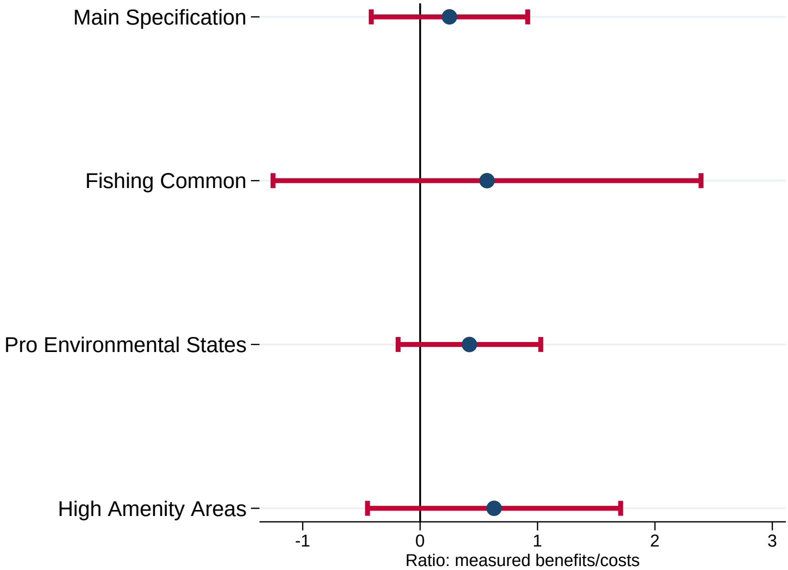
$$V_{pt} = \gamma G_{pt} + X_{pt}^{'} \beta + \eta_p + \eta_{wt} + \epsilon_{pt}$$

	(1)	(2)	(3)	(4)
Panel C. Log Total Housing Units				
Cumulative Grants	-0.007003**	-0.00046	-0.00041	-0.00018
	(0.003202)	(0.001186)	(0.000942)	(0.000242)
Panel D. Log Total Value of Housing Stock				
Cumulative Grants	-0.00253	0.00264	0.003594*	-0.00015
	(0.003952)	(0.002333)	(0.002095)	(0.000541)
Plant FE, Basin-by-Year FE	Yes	Yes	Yes	Yes
Dwelling Characteristics		Yes	Yes	Yes
Baseline Covariates * Year		Yes	Yes	Yes
Max Distance Homes to River: 0.25 Miles	Yes	Yes		
Max Distance Homes to River: 1.0 Miles			Yes	
Max Distance Homes to River: 25 Miles				Yes

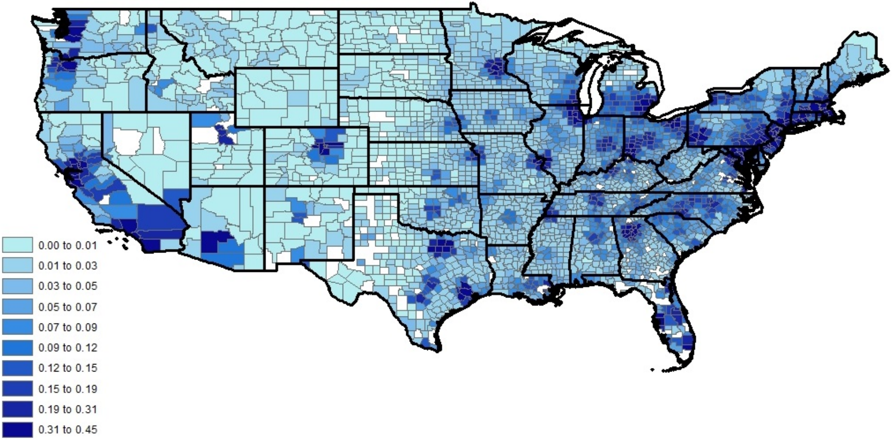
Change in Home Values and Costs

	(1)	(2)	(3)	(4)
Ratio: Change in Home	0.04	0.25	0.25	0.28
Values / Costs	(0.03)	(0.33)	(0.33)	(0.37)
p-value: Ratio = 0	[0.15]	[0.44]	[0.45]	[0.45]
p-Value: Ratio = 1	[0.00]	[0.02]	[0.02]	[0.05]
Change in Value of	13.06	86.12	85.03	108.07
Housing ($Bn)				
Costs ($Bn)				
Capital: Fed.	93.77	101.82	101.82	113.69
Capital: Local	38.56	41.64	41.64	47.81
Variable	180.53	196.45	196.45	221.82
Total	312.86	339.92	339.92	383.32
Max Distance Homes to River (Miles)	1	25	25	25
Include Rental Units			Yes	Yes
Include Non-Metro Areas				Yes

Ratio of Measured Benefits to Costs: Heterogeneity



Ratio of Measured Benefits to Costs: Heterogeneity



Interpreting Hedonic Estimates

Possible reasons to interpret estimates as lower bound on willingness to pay:

1. Incomplete information (health?)
2. May exclude some types of demand (non-use values?)
3. General equilibrium channels
    - Wages
    - Hedonic price function shifts
    - Substitution across sites
    - Ecology
4. 25 mile radius
5. Incomplete pass-through (crowd-out / displacement)?

Overview

- Background: water pollution and regulation
- Data
- Pollution trends
- Clean Water Act and pollution
- Clean Water Act and home prices
- **Conclusions**

Conclusions

- **Research Questions**
  - How has water pollution changed since 1972?
  - Did Clean Water Act causes changes?
  - How do residents value changes?

Conclusions

1. Pollution has declined substantially
2. Clean Water Act caused some of this
   - Costs \$1.5 million per year to make a river-mile fishable
   - Complete pass-through of federal grants to municipal spending
3. Change in home values $<$ costs
   - Overall ratio of measured benefits to costs: 0.25
   - Larger ratios ($\sim$0.6) in areas where fishing is popular
   - Interpret as lower bound?

Conclusions

- Estimated benefit-cost ratios for CWA much lower than for Clean Air Act.
- EPA: Clean Air Act benefit-cost ratio: 42. CWA: <1
- Why?
  1. No market based instruments
  2. Ignores major polluting sectors (agriculture)
  3. Easier to substitute away/defend from pollution
  4. Limited health consequences?

Conclusions

- Estimated benefit-cost ratios for CWA much lower than for Clean Air Act.
- EPA: Clean Air Act benefit-cost ratio: 42. CWA: <1
- Why?
  1. No market based instruments
  2. Ignores major polluting sectors (agriculture)
  3. Easier to substitute away/defend from pollution
  4. Limited health consequences?

- Ongoing related work:
  - The environmental bias of trade policy
  - Optimal pollution regulation for vehicles
  - Welfare consequences of industrial water pollution

Introduction: Health and Demand for Public Goods

- This is a big paper addressing several broad questions in very specific ways.
- How large is willingness to pay for health in rich v. poor countries?
  - Includes value of a statistical life and DALYs
- How should we estimate willingness to pay for health?
  - Revealed preference versus stated preference
- What property rights for public health investments maximize social welfare?

Introduction: Health and Demand for Public Goods

- This paper has three parts which could have been separate papers
  - But returns to paper quality are nonlinear, better to put in one paper

1. Experimental estimates: effect of water source improvement on water quality and health
2. Structural demand estimates: willingness to pay for water source improvement as revealed by travel behavior
   2.1 Compare to stated preferences
3. Property rights simulation

Background

- Unprotected springs: bacteria transmitted by hands
- Protected springs: encased in concrete
  - Little maintenance required
- 43% of Kenyans in the area use springs for drinking water
- Most springs on private land

Background

Experimental Design

- Spring Protection ($956 per spring)
- Implemented by local NGO, International Child Support
- Modest maintenance costs ($32/year)

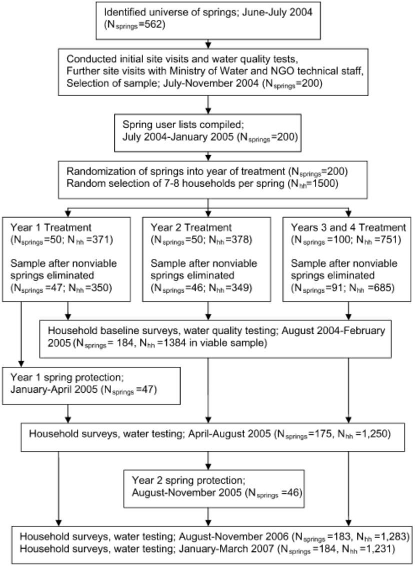
Experimental Design

- Obtain census of local unprotected springs from Ministry of Water and Irrigation
- Visit each spring, find ~575 suitable for protection
  - Randomly choose 200 to protect
- Compare springs improved in first year against springs chosen for later protection

Experimental Design

- Noncompliance: 16 springs in treatment and comparison groups later found unsuitable
  - 10 of remaining 184 springs had treatment differ from assignment
  - So use intent-to-treat estimates, also report IV/LATE (very similar)

Experimental Design



Experimental Design

TABLE I
BASELINE DESCRIPTIVE STATISTICS (2004 SURVEY)

	Treatment (protected)		comparison		Treatment – comparison
	Mean (s.d.)	Obs.	Mean (s.d.)	Obs.	(s.e.)
A: Spring-level data					
Ln. *E. coli* MPN (CFU/100 ml)	3.90	98	3.79	95	0.11
	(1.95)		(1.97)		(0.28)
Water is high quality (*E. coli* MPN ≤ 1)	0.05	98	0.06	95	–0.01
	(0.22)		(0.24)		(0.03)
Water is high or moderate quality (*E. coli* MPN < 126)	0.70	98	0.69	95	0.01
	(0.46)		(0.46)		(0.07)
Water is poor quality (*E. coli* MPN 126–1000)	0.19	98	0.23	95	–0.04
	(0.40)		(0.42)		(0.06)
Latrine density (fraction of homes with latrines)	0.85	98	0.88	95	–0.02
	(0.16)		(0.15)		(0.02)
Average diarrhea prevention knowledge score	3.06	98	3.19	95	–0.13
	(0.87)		(1.17)		(0.15)
Iron roof density (fraction of compounds with iron roof)	0.70	98	0.68	95	0.03
	(0.21)		(0.23)		(0.03)
Other variables used for randomization balancing					
Distance of spring from paved road (m)	3005	98	3028	95	–23
	(2101)		(2198)		(310)
Slope of catchment area (1 = flat, 5 = very steep)	3.56	98	3.59	95	–0.03
	(0.69)		(0.63)		(0.09)
Number of households that use the spring	29.90	98	29.60	95	0.30
	(13.99)		(14.33)		(2.04)

Experimental Design

TABLE I
(CONTINUED)

	Treatment (protected)		comparison		Treatment – comparison
	Mean (s.d.)	Obs.	Mean (s.d.)	Obs.	(s.e.)
	(13.99)		(14.33)		(2.04)
Butere district indicator	0.34	98	0.32	95	0.02
	(0.48)		(0.47)		(0.07)
Mumias district indicator	0.41	98	0.40	95	0.01
	(0.49)		(0.49)		(0.07)
Total coliform MPN (CFU/100 ml)	2170	98	2152	95	17
	(622)		(624)		(90)
E. coli MPN (CFU/100 ml)	265	98	248	95	17
	(548)		(552)		(79)
Water is poor or moderate quality (*E. coli* MPN 100–1000)	0.23	98	0.26	95	–0.03
	(0.43)		(0.44)		(0.06)
B: Household-level data					
Ln. *E. coli* MPN (CFU/100 ml)	3.22	733	3.33	712	–0.11
	(2.22)		(2.13)		(0.14)
Water is high quality (*E. coli* MPN ≤ 1)	0.15	733	0.12	712	0.04
	(0.36)		(0.32)		(0.02)**
Water is high or moderate quality (*E. coli* MPN < 126)	0.76	733	0.76	712	0.00
	(0.43)		(0.43)		(0.03)
Water is poor quality (*E. coli* MPN 126–1000)	0.17	733	0.16	712	0.01
	(0.37)		(0.37)		(0.02)

Experimental Design

TABLE I
(CONTINUED)

	Treatment (protected)		comparison		Treatment – comparison
	Mean (s.d.)	Obs.	Mean (s.d.)	Obs.	(s.e.)
Respondent years of education	5.71 (3.61)	731	5.66 (3.60)	717	0.05 (0.23)
Children under age 12 in the compound	4.04 (2.48)	736	3.93 (2.46)	719	0.10 (0.14)
Iron roof indicator	0.70 (0.46)	735	0.68 (0.47)	717	0.03 (0.03)
Walking distance to closest water source (minutes)	8.74 (8.40)	725	8.03 (6.82)	714	0.71 (0.49)
Water collection trips per week by household	48.03 (36.51)	733	47.99 (38.48)	716	0.04 (2.51)
Ever collects drinking water at "reference" spring indicator	0.82 (0.38)	661	0.80 (0.40)	668	0.02 (0.03)
Multisource user (uses sources other than reference spring)	0.45 (0.50)	732	0.44 (0.50)	715	0.00 (0.04)
Fraction of respondent water trips to "reference" spring	0.72 (0.41)	655	0.71 (0.42)	663	0.01 (0.04)
Rates water at the spring "very clean"—rainy season	0.33 (0.47)	736	0.33 (0.47)	719	0.00 (0.04)
Rates water at the spring "very clean"—dry season	0.74 (0.44)	736	0.74 (0.44)	719	–0.01 (0.03)
Fraction of water trips by those under age 12	0.10 (0.20)	727	0.10 (0.20)	711	–0.00 (0.01)

Experimental Design

TABLE I
(CONTINUED)

	Treatment (protected)		comparison		Treatment – comparison
	Mean (s.d.)	Obs.	Mean (s.d.)	Obs.	(s.e.)
Water storage container in home was covered	0.90	673	0.93	656	–0.03
	(0.30)		(0.26)		(0.02)**
Yesterday's drinking water was boiled indicator	0.25	731	0.29	711	–0.03
	(0.43)		(0.45)		(0.02)
Respondent diarrhea prevention knowledge score	3.06	736	3.19	719	–0.13
	(2.14)		(2.26)		(0.15)
Respondent said "dirty water" causes diarrhea	0.68	736	0.67	719	0.01
	(0.47)		(0.47)		(0.03)
Household has soap in the home	0.91	733	0.91	717	0.00
	(0.28)		(0.29)		(0.02)
C: Child demographics and health					
Child age (years)	1.70	1047	1.72	995	–0.02
	(0.95)		(0.97)		(0.04)
Child male (= 1)	0.52	1047	0.50	995	0.02
	(0.50)		(0.50)		(0.02)
Child had diarrhea in past week indicator	0.23	996	0.20	961	0.03
	(0.42)		(0.40)		(0.02)
Child height (cm)	76.10	870	76.13	835	–0.03
	(11.67)		(12.16)		(0.57)
Child weight (kg)	9.98	864	10.02	810	–0.05
	(3.04)		(3.09)		(0.16)

Notes. The treatment springs were later protected (in 2005). Huber–White robust standard errors are clustered at spring level when using household-level data, significant at * 90%, ** 95%, *** 99% confidence. Reference spring is based on spring user lists. Children in panel C were under age three at baseline or were born since then.

Spring Protection Impacts on Water Quality and Health

$$W_{jt}^{SP} = \alpha_t + \phi_1 T_{jt} + X_j^{SP'} \phi_2 + \left(T_{jt} * X_j^{SP} \right)' \phi_3 + \varepsilon_{jt}$$

- W_{jt}^{SP} water quality of spring j at time $t \in \{0, 1, 2, 3\}$
- T_{jt} treatment dummy
- X_j^{SP} baseline spring, community characteristics
- ε_{jt} error
- This is intent to treat estimate

Spring Protection Impacts on Water Quality and Health

TABLE II
SPRING PROTECTION SOURCE WATER QUALITY IMPACTS (2004–2007)

	Dependent variable ln(spring water *E. coli* MPN)			Water clarity (observed)	Water yield (observed)
	(1)	(2)	(3)	(4)	(5)
Treatment (protected) indicator	−1.07	−1.04	−1.10	0.26	−0.06
	(0.27)***	(0.23)***	(0.24)***	(0.07)***	(0.06)
Baseline ln(spring water *E. coli* MPN)		0.99	1.01		
		(0.07)***	(0.08)***		
Baseline ln(spring water *E. coli* MPN) *Treatment indicator		−0.17	−0.16		
		(0.12)	(0.13)		
Baseline latrine density			−0.07		
			(0.58)		
Baseline latrine density *Treatment indicator			0.90		
			(1.76)		
Baseline diarrhea prevention score			−0.04		
			(0.07)		
Baseline diarrhea prevention score *Treatment indicator			−0.29		
			(0.25)		
Baseline boiled water yesterday density			0.59		
			(0.68)		
Baseline boiled water yesterday density *Treatment indicator			0.92		
			(1.52)		
Baseline mother's years of education density			−0.06		
			(0.05)		

Spring Protection Impacts on Water Quality and Health

TABLE II

(CONTINUED)

	Dependent variable ln(spring water *E. coli* MPN) (1)	(2)	(3)	Water clarity (observed) (4)	Water yield (observed) (5)
Baseline mother's years of education density			0.06		
*Treatment indicator			(0.14)		
Treatment group 1 (phased in early 2005)			−0.25		
			(0.20)		
Treatment group 2 (phased in late 2005)			−0.17		
			(0.17)		
R^2	0.30	0.43	0.45	0.13	0.13
Observations	726	726	726	478	474
Mean (s.d.) of dependent variable in comparison group	3.63 (1.95)	3.63 (1.95)	3.63 (1.95)	0.76	0.80

Notes. Estimated using ordinary least squares. Huber–White robust standard errors are presented (clustered at the spring level), significantly different than 0 at *90%, **95%, *** 99% confidence. There are 184 spring clusters with data for some of the four survey rounds (2004, 2005, 2006, and 2007). MPN stands for "most probable number" colony-forming units (CFU) per 100 ml. Average diarrhea prevention knowledge calculated as average of demeaned sum of number of correct responses given to the open-ended question "to your knowledge, what can be done to prevent diarrhea?" Outcomes in columns 4 and 5 are enumerator assessments of spring water clarity and the spring's water yield. All variables that are interacted with the treatment indicator are de-meaned. Survey round and wave fixed effects are included in all regressions but not reported, as are all variables used to balance the initial randomization into treatment and comparison groups. Baseline iron roof density and its interaction with the treatment indicator (in column 3) are included as additional control variables (not shown in the table). The −1.07 effect in column 1 is equivalent to a 66% reduction in *E. coli* fecal coliform units per 100 ml.

Spring Protection Impacts on Water Quality and Health

TABLE III

SPRING PROTECTION HOUSEHOLD WATER QUALITY IMPACTS (2004–2007)

	Dependent variable ln(home water *E. coli* MPN)		
	(1)	(2)	(3)
Treatment (protected) indicator	−0.27	−0.29	−0.67
	(0.15)*	(0.19)	(0.27)**
Baseline ln(spring water *E. coli* MPN)	0.01	0.03	0.03
	(0.05)	(0.05)	(0.05)
Baseline multisource user		−0.29	−0.27
		(0.16)*	(0.17)
Baseline multisource user * Treatment indicator		0.04	0.06
		(0.25)	(0.26)
Baseline latrine density	−0.73	−0.73	−0.02
	(0.32)**	(0.31)**	(0.60)
Baseline latrine density * Treatment indicator			1.42
			(1.01)
Baseline diarrhea prevention score	−0.02	−0.03	−0.05
	(0.02)	(0.02)	(0.04)
Baseline diarrhea prevention score * Treatment indicator			−0.05
			(0.06)
Baseline boiled water yesterday indicator	0.17	0.16	0.29
	(0.08)**	(0.08)**	(0.15)*
Baseline boiled water yesterday indicator * Treatment indicator			0.52
			(0.28)*
Baseline mother's years of education	0.00	0.00	0.02
	(0.01)	(0.01)	(0.02)
Baseline mother's years of education * Treatment indicator			0.02
			(0.04)
Treatment group 1 (phased in early 2005)	0.00	−0.14	−0.01
	(0.14)	(0.18)	(0.27)
Treatment group 2 (phased in late 2005)	−0.10	−0.12	−0.16
	(0.12)	(0.15)	(0.27)
R^2	0.04	0.04	0.05
Observations (spring clusters)	4343 (184)	4343 (184)	4343 (184)
Mean (s.d.) of dependent variable in comparison group	3.00	3.00	3.00
	(2.27)	(2.27)	(2.27)

Notes. Estimated using ordinary least squares. Huber–White robust standard errors (clustered at the spring level) are presented, significantly different than 0 at * 90%, ** 95%, *** 99% confidence. MPN stands for "most probable number" colony-forming units (CFU) per 100 ml. Survey round and wave fixed effects included in all regressions but not reported, as are all variables used to balance the initial randomization into treatment and comparison groups. Additional control variables are: number of children under 12 living in the home, home has iron roof indicator, iron roof density within spring community. When differential treatment effects are reported in column (3), we also include interactions of these control variables with the treatment (protected) indicator (not shown in the table). Baseline spring water quality, latrine density, diarrhea prevention score, and mother's education are de-meaned. The −0.27 effect in column 1 is equivalent to a 24% reduction in *E. coli* fecal coliform units per 100 ml.

Spring Protection Impacts on Water Quality and Health

For child health and anthropometrics, use

$$Y_{ijt} = \alpha_i + \alpha_t + \phi_1 T_{jt} + X_{ij'}\phi_2 + (T_{jt}X_{ij})\phi_3 + u_{ij} + \varepsilon_{ijt}$$

Child i
Time t
Spring j

Spring Protection Impacts on Water Quality and Health

TABLE IV
HEALTH OUTCOMES FOR CHILDREN UNDER AGE THREE AT BASELINE OR BORN SINCE 2004 (2004–2007 DATA)

	Dependent variable: diarrhea in past week						Dependent variable: weight (kg)		Dependent variable: BMI (kg/m²)	
	(1)	(2) Probit	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment (protected) indicator	−0.045***	−0.044**	−0.045*	−0.047**	−0.090***	−0.032	0.065	0.093	0.21	0.27
	(0.012)	(0.021)	(0.023)	(0.023)	(0.029)	(0.039)	(0.075)	(0.100)	(0.13)*	(0.16)
Treatment (protected) indicator * Male					0.083**			−0.054		−0.12
					(0.040)			(0.121)		(0.18)
Treatment (protected) indicator * Baseline latrine density						0.105				
						(0.123)				
Treatment (protected) indicator * Baseline diarrhea prevention score						−0.008				
						(0.007)				
Treatment (protected) indicator * Baseline mother's years of education						0.002				
						(0.004)				

Spring Protection Impacts on Water Quality and Health

TABLE IV
(CONTINUED)

	Dependent variable: diarrhea in past week						Dependent variable: weight (kg)		Dependent variable: BMI (kg/m^2)	
	(1)	(2) Probit	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Child fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Treatment group fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month of year controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender-age controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.00	—	0.53	0.53	0.53	0.53	0.96	0.96	0.69	0.69
Child-year observations	6750	6749	6749	6660	6660	6601	5736	5736	5646	5646
Mean (s.d.) of the dependent	0.19	0.19	0.19	0.19	0.19	0.19	11.36	11.36	17.0	17.0
variable in the comparison group	(0.39)	(0.39)	(0.39)	(0.39)	(0.39)	(0.39)	(3.50)	(3.50)	(2.2)	(2.2)

Notes. Column (2) estimated using probit (marginal effects presented), columns (1) and (3)–(10) estimated using ordinary least squares. Huber-White robust standard errors (clustered at the spring level) are presented, significantly different than 0 at * 90%, ** 95%, *** 99% confidence. Data from all four survey rounds (2004, 2005, 2006, 2007), sample restricted to children under age three at baseline (in 2004) and children born since 2004. Diarrhea defined as three or more "looser than normal" stools within 24 hours at any time in the past week. The gender-age controls include linear and quadratic current age (by month), and these terms interacted with a gender indicator. Columns (3)–(10) also contain survey round controls. In column (6), additional control variables are number of children under 12 living in the home, home has iron roof indicator, iron roof density within spring community, and the boiled water yesterday indicator (all measured at baseline), all interacted with the treatment indicator.

Spring Protection Impacts on Water Quality and Health

TABLE V
Treatment Effects on Household Water Source Choice and Health Behaviors (2004–2007)

Dependent variable	Coefficient (s.e.) on treatment indicator full sample (1)	Coefficient (s.e.) on treatment indicator sole-source users (2)	Coefficient (s.e.) on treatment indicator multisource users (3)	Mean (s.d.) comparison group in 2006, 2007 surveys (4)
A: Water transportation and storage				
Fraction of water trips by those under age 12[a]	0.00 (0.01)	0.00 (0.02)	0.00 (0.02)	0.09 (0.19)
Water storage container in home covered indicator	0.00 (0.01)	−0.01 (0.02)	0.01 (0.02)	0.98 (0.15)
Ever treated water with chlorine indicator[b]	0.02 (0.03)	0.03 (0.05)	0.01 (0.04)	0.45 (0.50)
Yesterday's drinking water boiled indicator[c]	0.03 (0.02)	0.05 (0.03)*	0.01 (0.03)	0.25 (0.44)
B: Sanitation and hygiene behaviors				
Diarrhea prevention knowledge score	0.14 (0.14)	0.21 (0.18)	0.04 (0.19)	3.92 (2.07)
Respondent says drinking clean water is a way to prevent diarrhea	−0.03 (0.03)	−0.03 (0.04)	−0.04 (0.04)	0.73 (0.44)
Household has soap in the home indicator	−0.01 (0.02)	−0.02 (0.02)	0.01 (0.03)	0.89 (0.31)
Fingers with bacterial contamination (fecal *Streptococci* colonies)[d]	0.10 (0.12)	0.41 (0.23)*	0.11 (0.21)	0.71 (1.26)

Spring Protection Impacts on Water Quality and Health

TABLE V
(CONTINUED)

Dependent variable	Coefficient (s.e.) on treatment indicator full sample (1)	Coefficient (s.e.) on treatment indicator sole-source users (2)	Coefficient (s.e.) on treatment indicator multisource users (3)	Mean (s.d.) comparison group in 2006, 2007 surveys (4)
C: Water collection and source choice				
Fraction of trips to reference spring	0.09 (0.03)***	0.03 (0.02)*	0.21 (0.05)***	0.76 (0.40)
Perceive water at reference spring to be very clean—rainy season	0.22 (0.04)***	0.22 (0.05)***	0.22 (0.04)***	0.18 (0.38)
Perceive water at reference spring to be very clean—dry season	0.11 (0.04)***	0.07 (0.03)**	0.15 (0.06)***	0.76 (0.43)
Trips made to get water (all uses, members, sources) past week	−2.38 (2.15)	−0.71 (2.41)	−4.41 (3.51)	31.77 (24.42)

Notes. $N = 1354$ households at 184 springs (full sample), 755 of whom are baseline sole source users. Each cell reports the differences-in-differences treatment effect estimate from a separate regression, where the dependent variable is reported in the first column. Huber–White robust standard errors (clustered at the spring level) are presented, significantly different than 0 at * 90%, ** 95%, and *** 99% confidence. Reported means of the dependent variables are in the comparison group 2006 and 2007 (rounds 2 and 3 post-treatment) surveys. Reference spring is the sample spring that we believed households used at baseline based on spring user lists. The fingertip contamination results are for the respondent's main hand (values range from 0 to 5).

[a] Because of changes in survey design, responses to this question are not available for the third (2006) round of data collection.

[b] Because of changes in survey design, responses to this question are not available for the first (2004) round of data collection.

[c] Because of changes in survey design, responses to this question are not available for the fourth (2007) round of data collection.

[d] Because information on fingertip contamination was collected only in the third (2006) round of data collection, this cell reports the difference between the treatment and comparison groups rather than the differences-in-differences treatment effect.

Demand Estimates

$$u_{ijt} = \beta_i T_{jt} + Z_j - C_i D_{ij} + e_{ijt}$$

u_{ijt} Indirect utility: household i, source j, time t
Z_j Value of water from source j
T_{jt} Treatment
β_i Effect of treatment on indirect utility
$C_i > 0$ Household i cost of time per minute
D_{ij} Household's round-trip distance to source j
e_{ijt} IID type I extreme value error

Demand Estimates

Household i chooses source j over k if

$$\beta_i \left(T_{jt} - T_{kt} \right) + \left(Z_j - Z_k \right) - C_i \left(D_{ij} - D_{ik} \right) + \left(e_{ijt} - e_{ikt} \right) \geq 0$$

Given characteristics X_{ijt} for household, source, time, choice probabilities have logit form:

$$\Pr \left(y_{ijt} | X \right) = \frac{\exp \left(X'_{ijt} B \right)}{\sum_h \exp \left(X'_{iht} B \right)} \equiv \rho_{ijt}$$

$h \in H$ alternatives
$y_{ijt} = 1$ Dummy: household i chooses j at time t
Mixed logit has random coefficients on spring protection and walking distance:

$$\Pr \left(y_{ijt} | X \right) = \int_B \rho_{ijt} f \left(B \right) dB$$

Bayesian numerical methods to maximize log-likelihood

Demand Estimates

TABLE VI
DISCRETE CHOICE MODELS (CONDITIONAL AND MIXED LOGIT) OF WATER SOURCE CHOICE (2007 SURVEYS)

	Revealed preference					Stated ranking	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment (protected) indicator	0.51***	−0.02	0.57***	0.68***		0.96***	
	(0.051)	(0.08)	(0.08)	(0.09)		(0.24)	
Mixed logit—mean (normal):					2.95***		1.46**
					(0.25)		(0.60)
Mixed logit—s.d. (normal):					5.73***		1.22
					(0.33)		(0.75)
ln(source water *E. coli* MPN)		−0.14***					
		(0.01)					
Water quality at source perceived to be above average		1.14***					
		(0.07)					
Distance to water source (minutes walking)	−0.055***	−0.059***	−0.047***	−0.053***		−0.033***	
	(0.002)	(0.002)	(0.003)	(0.002)		(0.010)	
Mixed logit—mean (restricted triangular):					−0.21***		−0.03***
					(0.01)		(0.01)
Mixed logit—s.d. (restricted triangular)					0.09		0.01

Demand Estimates

TABLE VI
(CONTINUED)

	Revealed preference					Stated ranking	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Distance * children aged 0–3			–0.008***				
			(0.002)				
Treatment indicator * children aged 0–3			0.04				
			(0.04)				
Treatment indicator * baseline latrine ownership				1.80***			
				(0.25)			
Treatment indicator * baseline diarrhea prevention score				0.023			
				(0.020)			
Treatment indicator * baseline mother's years of education				0.058***			
				(0.011)			
Source type: borehole/piped	–0.08		–0.10	–0.13*	–1.02***	0.07	0.04
	(0.07)		(0.07)	(0.08)	(0.14)	(0.25)	(0.27)
Source type: well	–0.28***		–0.31***	–0.31***	–1.87***	–0.43*	–0.47*
	(0.06)		(0.07)	(0.07)	(0.13)	(0.24)	(0.25)
Source type: stream/river	–0.77***		–0.70***	–0.63***	–1.46***	–2.19***	–2.25***
	(0.09)		(0.09)	(0.09)	(0.15)	(0.52)	(0.53)

Demand Estimates

TABLE VI
(CONTINUED)

	Revealed preference					Stated ranking	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Source type: lake/pond	–0.20		–0.30	–0.18	–0.32	–2.82	–2.85
	(0.20)		(0.20)	(0.19)	(0.35)	(1.86)	(1.87)
Number of observations (water collection choice situations)	53427	29068	50988	50024	53427	2114	2114
Number of households	452	329	428	422	452	483	483

Notes. The data are from the final round of household surveys (2007). Conditional logit in columns (1)–(4) and (6), and mixed logit in columns (5) and (7) (grouped by choice and weighting households equally). Significant at * 90%, ** 95%, and *** 99% confidence. In columns (1)–(5) each observation is a unique household-water source pair in one water collection trip. In columns (6)–(7), each observation is a household-water source pair from questions where the respondent chooses their preferred source. The dependent variable is an indicator equaling 1 if the household chose the water source represented in the household-source pair. The omitted water source category is "non-program spring". The coefficient estimate on the indicator for the household's reference sample spring is included in the analysis but not shown in the table. In column (3), additional controls are included for children aged 3–12 at baseline interacted with the treatment indicator and distance to the water source (not shown). In column (4), additional controls are the number of children under 12, home has iron roof indicator, iron roof density in the community, and the boiled water yesterday indicator (all measured at baseline), directly and interacted with the treatment indicator.

Demand Estimates

- Concern about measurement error and attenuation bias in reported walking distance
  - Correlation across survey rounds is 0.38
  - So in applying the estimates, inflate coefficient by -0.055/0.38=-0.145
- Could have used better methods if knew this issue ex ante?

Demand Estimates

- Two stated preference methods
  - Ask households to rank water sources
  - Ask households hypothetical willingness to pay ("contingent valuation")

Demand Estimates

TABLE VII
VALUATION OF ONE YEAR OF SPRING PROTECTION (2007 SURVEY)

	One year of spring protection	
	Mean	s.d.
A: Revealed preference valuation (from mixed logit; Table VI, column 5)		
Work days (8-hour days)	32.4 days	102.8 days
Time value from survey questions (time and monetary value)	$2.96	$11.14
Assume value of time is 25% Kenyan worker average wage	$11.57	$36.69
B: Stated preference ranking valuation (from mixed logit; Table VI, column 7)		
Work days (8-hour days)	56.2 days	12.3 days
Time value from survey questions (time and monetary value)	$4.96	$1.97
Assume value of time is 25% Kenyan worker average wage	$20.06	$4.38
C: Contingent valuation	Full round	Final wave, emphasizing trade-offs
Proportion willing to pay this for spring protection:		
US$3.57 (250 Kenyan shillings)	0.94 [308]	0.80 [98]
US$7.14 (500 Kenyan shillings)	0.90 [316]	0.79 [204]
US$14.29 (1000 Kenyan shillings)	—	0.60 [204]
	One year of spring protection	
	Mean	s.d.
Sample: Final wave, emphasizing trade-offs	$17.64	$13.09
Subsample with 250 KSH starting value	$12.62	$11.06
Subsample with 500 KSH starting value	$23.91	$14.28

Notes. The number of observations is in brackets in C. The contingent valuation questions were only asked of households in the treatment group, since they have firsthand knowledge of protection. In the final wave of the survey, respondents were first asked if they would be willing to pay either 500 or 1000 Kenyan shillings, followed by the question that emphasized the expenditure trade-off for their assigned amount, and then were asked if they would be willing to pay the next higher amount, also with emphasis on the expenditure trade-off.

Demand Estimates

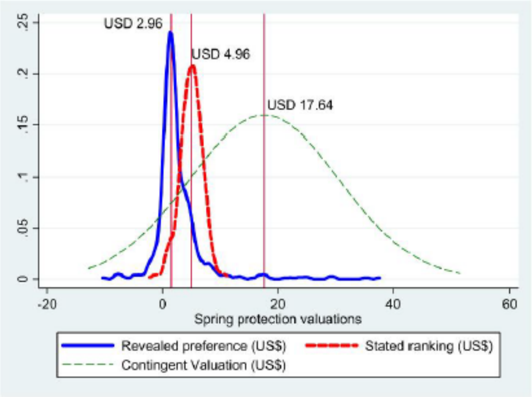


FIGURE II
Household Revealed Preference and Stated Preference Valuations of One Year of Spring Protection (2007)

The revealed preference estimates are from the mixed logit results in Table VI, regression 5, and the stated preference ranking results are from the mixed logit results in Table VI, regression 7. The contingent valuation data are presented in Table VII, panel C.

Demand Estimates

Implications for health demand

- Spring protection prevents 3.2 diarrhea cases per household-year
- Mixed logit implies spring protection has value of 32.4 work days
- So value of statistical life is 8,742 work days or 35 work years
- In monetary terms,
  - Demand for avoiding a case of child diarrhea: US$0.89
  - Demand for avoiding a child diarrhea death: $769
    - With greater value of time, rises to $83.61 per DALY, $2,715 per diarrhea death
  - Upper bound on DALY: $23.68

Demand Estimates

- The implied willingness-to-pay for health is tiny!
  - World Bank 1993 World Development Report: policies costing less than \$150/DALY are "extremely cost effective"
  - Other standard: \$100/DALY
  - Sachs: Kenya DALY is >\$400 based on relative GDP per capita
  - VSL in US is \$2 to \$7 million
  - Hall and Jones: high income elasticity of demand for health, so Kenya VSL is \$477 to \$1,603 (reasonable)

Alternative Property Rights

- Assume marginal cost of water is zero (no pump, user congestion, or cost to unused water)
- Game timing:
  - s=0: property rights regime chosen
  - s=1: spring owners choose whether to protect
  - s=2: spring owners simultaneously set water prices with full knowledge
  - s=3: households choose water sources to maximize utility given protection decisions and prices

Alternative Property Rights

- Solve model backwards to find Nash equilibrium
  - [Does an equilibrium exist, is it unique?]
  - Consider subgroups of ≤ 4 springs within a sublocation
  - Assume resale permitted (implies linear pricing)
  - Search for fixed point of best response functions

Alternative Property Rights

- Models of property rights
  - Freehold (owner decides)
  - Lockean: only charge if improve
  - Modified Lockean: can charge for protected, not unprotected
  - Public funding with 30% DWL taxation
  - Public funding of vouchers

Alternative Property Rights

TABLE VIII
PROPERTY RIGHTS NORMS AND INSTITUTIONS: COUNTERFACTUAL SIMULATIONS

	Springs protected (%)	Average price per water trip (USD) | price > 0	Spring owner profits (USD), per spring	Household utility (USD) per spring	Average walking time (min)	Average fecal contamination, ln (avg *E. coli*)	Social welfare (USD) per spring
A: Revealed preference valuation, neoclassical planner							
Communal property (status quo)	0	0	0	0	11.6	4.66	0
Social planner	29	0	0	722	12.5	4.44	349
Freehold private property	5	0.0027	417	–508	12.9	4.92	–91
Lockean private property	12	0.0069	83	–127	12.3	4.70	–43
Modified Lockean property	2	0.0058	8	26	11.7	4.66	34
Public investment	22	0	0	507	12.4	4.48	130
Public vouchers	11	0.0012	72	336	12.1	4.56	124

Alternative Property Rights

TABLE VIII
(CONTINUED)

	Springs protected (%)	Average price per water trip (USD) | price > 0	Spring owner profits (USD), per spring	Household utility (USD) per spring	Average walking time (min)	Average fecal contamination, ln(avg *E. coli*)	Social welfare (USD) per spring
B: Paternalistic social planner							
Communal property (status quo)	0	0	0	0	11.6	4.66	0
Social planner	74	0	0	2,412	12.7	4.31	2,540
Freehold private property	5	0.0027	417	–508	12.9	4.92	—[a]
Lockean private property	12	0.0069	83	–127	12.3	4.70	—[a]
Modified Lockean property	2	0.0058	8	26	11.7	4.66	91
Public investment	71	0	0	929	12.7	4.32	2,113
Public vouchers	46	0.0032	765	775	12.4	4.43	1,732

Notes. [a]Social welfare in the freehold and Lockean cases in B cannot be reliably compared to the communal property case, as described in Section IV.C.
Profits, utility, and welfare are net present values (5% annual discount over 15 years). Household spring protection valuations are from Table VI, column 5, and utility is converted into USD using households' predicted time value. The Neoclassical planner values spring protection at households' revealed preference level, while the paternalistic planner values it at US$125/DALY averted. A summary of key assumptions is as follows. Communal property rights: the price of spring water is 0. No springs are protected. Social welfare is normalized to 0. Social planner: planner maximizes social welfare. The price of spring water is 0, its marginal cost. There is no deadweight loss to raising funds for spring protection. The planner knows preferences θ_{ij} (protection valuation, disutility of walking time) for each household. Freehold private property rights: spring owners simultaneously choose whether to protect springs and then simultaneously choose price per unit of water noncooperatively (in groups of up to four). Spring owners know preferences θ_{ij} for each household. Lockean private property rights: same as the freehold private case except the price of unprotected spring water is constrained to be 0. Modified Lockean property rights: same as the Lockean private case except the spring owner must always provide access to free unprotected water. Public investment: policy maker maximizes social welfare. The price of spring water is 0. There is 30% deadweight loss to raising funds for spring protection. The policy maker knows the distribution of preferences $F(\theta_{ij})$ in the population but not preferences for each household. Vouchers: the policy maker sets the voucher price for protected spring water to maximize social welfare, taking into account effects on spring owners' subsequent investment. Spring owners then make profit-maximizing protection decisions in simultaneous noncooperative play. There is 30% deadweight loss to funding the vouchers. The policy maker knows the distribution of preferences $F(\theta_{ij})$ in the population; spring owners know preferences θ_{ij} for each household.

Conclusions

- This is a big paper (in many respects)
- Nice example of using an experiment to estimate primitive utility parameters
- Like many good papers, pertinent to several fields: development, health, environment, PF, IO, ...

Introduction

- Typical paper: estimate a model with confidence regions reflecting sampling variation in the data
- But in many settings economics agents are uncertain about the underlying parameters and learn about them over time
- In some cases the econometrician observes behavior and infers how it reflects learning versus stable policy functions
- In others we decide policy functions ex ante
- General question: how does learning modify benefit-cost analysis?

Introduction: Gallagher

- Background on how this paper came about

Introduction: Gallagher

- Floods: rare and potential catastrophic event
  - Parcel-level maps available but homeowners may ignore
- This paper: new data on flooding and insurance purchase
  - From FOIA (Freedom of Information Act request)
- Record all flood insurance policies purchased
- Record when each community is hit by Presidential Disaster Declaration Flood (PDD)
  - 18 years
  - 27 million insurance policies
  - 11,025 county flood
  - 643 PDD floods
  - 92 percent of counties hit by $>=$1 flood

Introduction: Gallagher

- Insurance policies per capita measure beliefs
  - Insurance prices don't change before v. after a flood.
- Assume community-level flood probability is constant 1958-2007

Findings

- Insurance purchase jumps after flood, peaking at 9 percent increase
- Takes nine years for effect to finish.
- A flood provides new information given historical data. Why do purchases jump?
  - Homeowners forget about past floods
  - Migration if people only use personal experience in the community
- Communities nearby and in same TV market also affected, albeit less
- Basic Bayesian learning model matches takeup but not decline
  - Need learning model with weight on more recent experience
  - Behavioral economics explanation: Availability Bias
- Suggestive role for migration

Background: Flood Insurance

- Until 1960s, limited flood insurance
- 1968: National Flood Insurance Program (NFIP)
- Choose actuarially fair prices, then markup 30-40 percent to cover expenses
- Prices based on historical flood data, hydrological models, community flood maps from Army Corps of Engineers
- NFIP specifies ten flood zones
  - Highest risk: 100-year flood plain
- Homeowners choose each year whether to buy insurance
  - Private firms sell at NFIP-specific rates
  - Premium must be paid each year
  - Flood insurance required at home purchase in 100-year flood plain

Background: Flood Insurance

- Premium doesn't change if a home is flooded
  - Real prices almost constant over this period
- FOIA request: NFIP data on all policies, 1980-2007
  - Number of policies grew 2 million in 1980 to 5.5 million 2007.
  - This paper doesn't explain broad growth

Background: Presidential Disaster Declaration Floods

- Created from 1950 Disaster Relief Act
  - Governor writes the U.S. president requesting funds for specific counties
  - Based on disaster scope, damage estimates
  - 75% of requests are approved
- Two types of disaster assistance
  - Public Assistance: governments and NGOs get funds to remove debris, repair infrastructure, reconstruct buildings
  - Individual Assistance: low-interest loans for rebuilding
- Data cover PDD events. Date, disaster type, county, estimate of disaster cost.
- Covers 1990-2007
- Covers 86 percent of US counties. 92 percent of them hit by at least one flood in 1990-2007
  - Median county hit by three floods.

Background: Presidential Disaster Declaration Floods

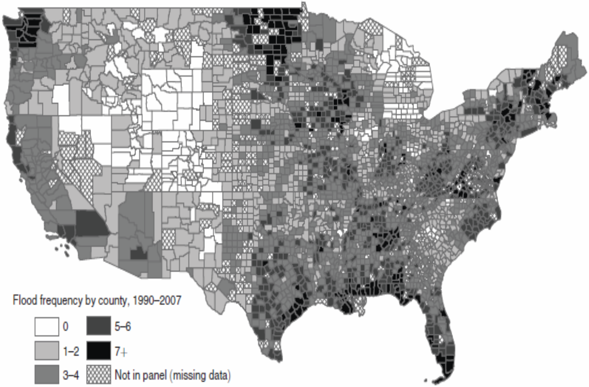


FIGURE 1. PRESIDENTIAL DISASTER DECLARATION FLOOD INTENSITY BY COUNTY 1990–2007

Notes: Map of the continental United States delineated by county. The map shows the Presidential Disaster Declaration (PDD) flood intensity by county from 1990–2007 for the counties included in the 1990–2007 community panel. Counties with no coloring have zero PDD floods. The darker the shade of grey, the greater the number of PDD floods. Counties with hash marks are not included in the panel. A county is dropped from the panel if no community in the county is included in the 1990–2007 community panel. The online text and data Appendix provide more details on the Presidential Disaster Declaration flood data and the 1990–2007 community panel.

Background: Presidential Disaster Declaration Floods

- Separate FOIA request: every Public Assistance damage claim paid, 1990-2007
  - 800,000 observations
  - Indicator for whether communicty within PDD is hit by a flood
  - Community is "hit" if $>=1$ claim in the community
  - 32% of communities in PDD counties are hit in the year of a flood
- Assume flood probabilities constant over 1958-2007.
  - This is claim of Army Corps of Engineers and NFIP
  - Few maps have been updated since they were created in 1970s and 1980s.
- Assume no serial correlation in floods
  - Test with Wald-Wolfowitz Runs Test

Econometric Model

- Event Study Regression:

$$\ln\left(takeup_{ct}\right) = \sum_{\tau=-T}^{T} \beta_\tau W_{c\tau} + \alpha_c + \gamma_{st} + \varepsilon_{ct}$$

- Observation: community-by-year
- Community: defined by FEMA, roughly equal to a Place (village/town/city/etc.)
- Takeup: flood policies per person
- W_{c0}: year of a flood
- Some communities hit by multiple floods in nearby years
  - Allow several event study variables to equal one for the various floods
  - Example: a community is hit by PDD flood in 1991 and 2004. In 2000, $W_{c9} = 1$ and $W_{c-4} = 1$.
- Combine outlying bins into one
  - They're identified from few observations
  - Combining increases statistical power
- Standard errors clustered at state level

Econometric Model

- Separate regressions: allow effects for nearby communities

$$\ln\left(takeup_{ct}\right) = \sum_{\tau=-T}^{T} \beta_\tau W_{c\tau} + \sum_{\tau=-T}^{T} \lambda_\tau N_{c\tau} + \alpha_c + \gamma_{st} + \varepsilon_{ct}$$

- $N_{c\tau}$ neighbors not hit by floods
- Two samples: 1980-2007; 1990-2007
  - Later sample: distinguishes communities hit within a county
- Advantages of 1980-2007 panel:
  - Control for lagged PDD floods from before start of panel
    - County data from earlier, not community data
  - County data is used to define neighbors
  - National migration and income data available at county level
  - County data nice for learning model since historic floods observed at county level.

Results

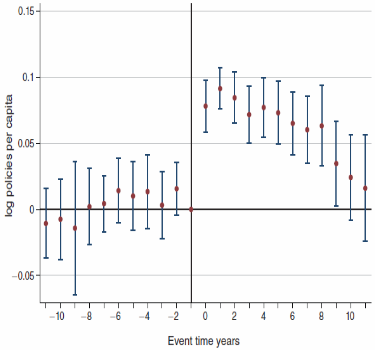
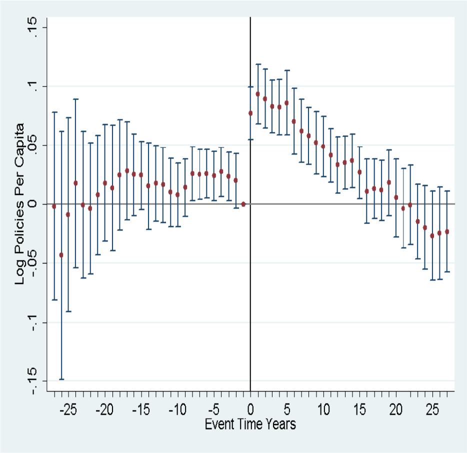


Figure 2. Flood Insurance Take-Up for Communities Hit by a Presidential Disaster Declaration Flood 1990–2007

Notes: The figure plots event time insurance take-up coefficients from estimation of equation (1) on the 1990–2007 panel. All estimated coefficients can be interpreted as the percent increase in flood insurance policies per capita for a hit community relative to the year before a flood (−1 on the *x*-axis). The end points on the graph are binned so that −11 (+11) is a bin for years −11 to −17 (+11 to +17). The vertical axis measures log per capita flood insurance take-up. The coefficient for the year before a flood is normalized to zero. The bars show the 95 percent confidence interval. Standard errors are clustered by state. There are 10,841 communities in the event study. A community is defined by the National Flood Insurance Program (NFIP) and corresponds to political jurisdictions: city, town, village, etc. A community is defined as hit if there is a Public Assistance damage claim submitted to the Federal Emergency Management Agency (FEMA) for damage from a Presidential Disaster Declaration flood.

Results

Figure 1: Community Flood Insurance Take-up 1980-2007 Panel



This figure shows insurance take-up point estimates from the same specification as Figure 4 in the text, except that the end points of the event study are not binned. The figure plots event study coefficients from the estimation of Equation (1) using the 1980-2007 panel. The designation of a flood is whether the community is located in a Presidential Disaster Declaration County. Refer to Section 3 of the text and the notes to Figure 4 for more details.

Results

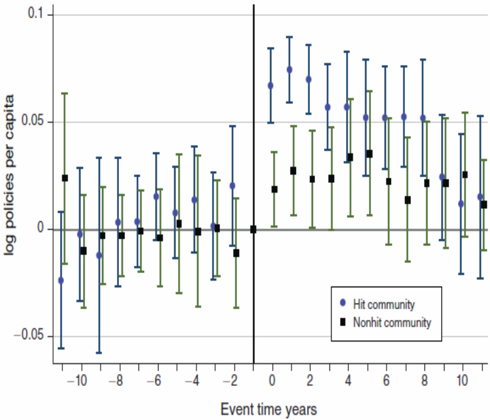


Figure 3. Flood Insurance Take-Up for Hit and Nonhit Communities within Presidential Disaster Declaration Flooded Counties 1990–2007

Notes: The figure plots event time insurance take-up coefficients from estimation of equation (2) on the 1990–2007 panel. The event study specification includes a set of indicators for nonhit communities within PDD counties. Please refer to the notes to Figure 2 and to Section III for more details on the event study specification.

Results

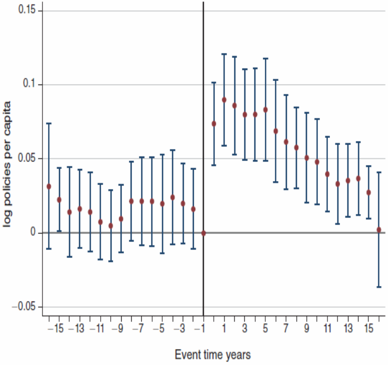


Figure 4. Flood Insurance Take-Up for Communities in Presidential Disaster Declaration Counties 1980–2007

Notes: The figure plots event time insurance take-up coefficients from estimation of equation (1) on the 1980–2007 panel. All estimated coefficients can be interpreted as the percent increase in flood insurance policies per capita for a hit community relative to the year before a flood (−1 on the *x*-axis). The end points on the graph are binned so that −16 (+16) is a bin for years −16 to −27 (+16 to +27). The vertical axis measures log per capita flood insurance take-up. The coefficient for the year before a flood is normalized to zero. The bars show the 95 percent confidence interval. Standard errors are clustered by state. There are 9,607 communities in the event study. A community is defined by the National Flood Insurance Program (NFIP) and corresponds to political jurisdictions: city, town, village, etc. A community is defined as hit if it is in a PDD flooded county.

Results

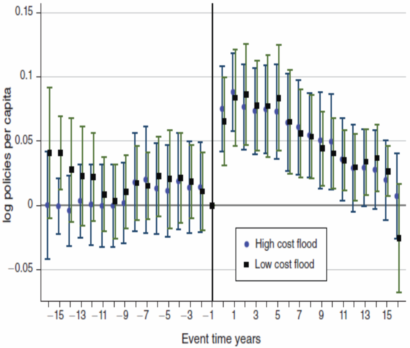


FIGURE 5. FLOOD INSURANCE TAKE-UP FOR COMMUNITIES IN PDD COUNTIES AFTER HIGH- AND LOW- COST FLOODS 1980–2007

Notes: The figure plots event time insurance take-up coefficients from estimation of a version of equation (1) that separately identifies floods as above (circles) or below (squares) per capita median cost on the 1980–2007 panel. Per capita cost is calculated over all 836 floods from 1980–2007 by dividing (a measure of) total PDD cost by the total population living in the effected counties in the year of a flood. The per capita cost ranges from less than \$1 to \$12,440, with a mean of \$70, and a median of \$20. Costs include all Public Assistance and Individual Assistance paid out after a flood. Please refer to the notes to Figure 4 for more details on the event study, and online Appendix Section D for a detailed cost data description.

Source: Public Entity Risk Institute

Results

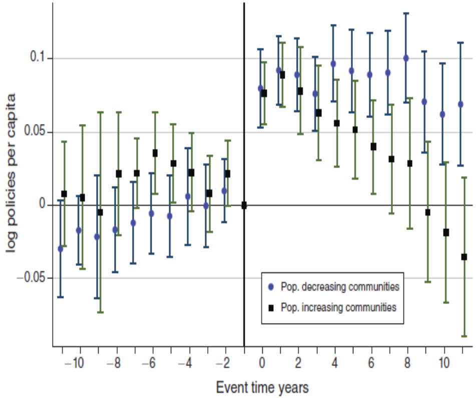
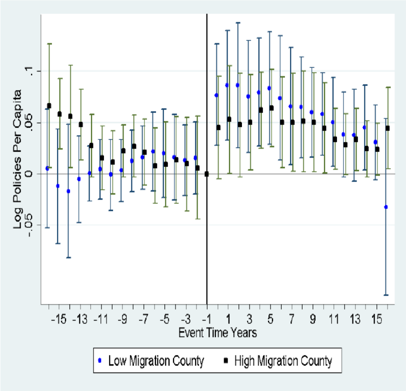


FIGURE 6. FLOOD INSURANCE TAKE-UP FOR POPULATION INCREASING AND POPULATION DECREASING COMMUNITIES 1990–2007

Notes: The figure plots event time insurance take-up coefficients from estimation of a version of equation (1) that separately identifies floods that hit population increasing (circles) and decreasing (squares) communities on the 1990–2007 panel. A community is defined as having an increasing (decreasing) population if its population grew (shrank) between 1990 and 2007. 6,113 (56 percent) of the 10,841 communities have a growing population. Refer to the notes to Figure 2 for more regression details.

Results

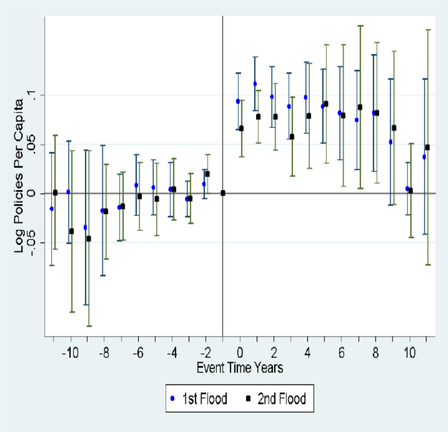
Figure 2: Flood Insurance Take-up after a Flood for Homeowners in Low and High Migration Counties, 1984-2007 Panel



The figure plots event time insurance take-up coefficients from the estimation of a version of Equation (1) that separately measures the effect on homeowners living in low (circles) and high (squares) migration counties. The bars show the 95% confidence intervals. There are two differences between this specification and the baseline estimation of Equation (1) using the 1980-2007 Panel shown in Figure 4 of the text. First, two sets of event time indicators are included in the specification to separately measure the effect for homeowners in above and below migration counties. A low (high) migration county is one that is below (above) median among all counties in our sample based on the average county migration rate from 1984-2007. Two sets of event time indicators are included in the specification to separately measure the effect for homeowners in above and below migration counties. Second, county migration data are only continuously available for the years 1984-2007, so the panel is restricted to the years 1984-2007.

Results

Figure 3: Flood Insurance Take-up after 1st and 2nd Flood Hits, 1990-2007 Panel



This figure plots event time insurance take-up coefficients from the estimation of a version of Equation (1) that separately measures the effect of 1st (circles) and 2nd (squares) flood hits using the 1990-2007 Panel. The bars show the 95% confidence intervals. The baseline estimation of Equation (1), shown in Figure 2 of the text, provides the average effect across all floods. There are 10,841 communities in the panel. 6,914 communities are hit by at least one flood and 3,822 by at least two floods. Refer to Section 3 of the text and the notes to Figure 4 for more details.

Results

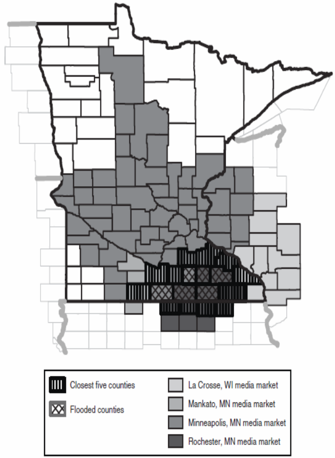


FIGURE 7. TV MEDIA MARKET AND GEOGRAPHIC NEIGHBOR MINNESOTA IDENTIFICATION EXAMPLE

Notes: The figure shows the state of Minnesota outlined in black. In 2004, six counties (marked by crossing lines) in Minnesota had a PDD flood. The parallel vertical lines indicate counties that are among the five closest counties to a flooded county and also not flooded. The flooded and (five closest) geographic neighbor counties are part of four different media markets. Counties in the four media markets are denoted by shades of grey. Closest counties are determined by Euclidean distance between county centroids. Nielson Media Research classifies each US county as belonging to a primary television media market. Please refer to Section IIIB for details.

Results

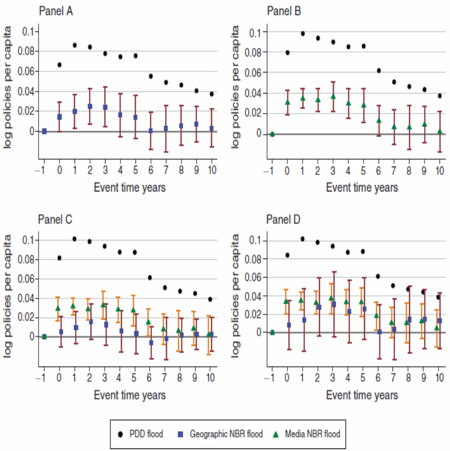


FIGURE 8. FLOOD INSURANCE TAKE-UP FOR GEOGRAPHIC AND MEDIA NEIGHBORS

Notes: Each panel contains coefficients from a distinct event study regression using a version of equation (2) and the 1980–2007 panel. Panel A includes event time indicators for communities located in one of the five closest nonflooded counties. Panel B includes event time indicators for nonflooded communities located in the same TV media market as a flooded community. Panel C includes both geographic and media indicators. Panel D includes both geographic and media indicators, and their interaction (not displayed). Please refer to Section IIIB and online Appendix Section D for further details.

Results

TABLE 1—TV MEDIA MARKET AND PRESIDENTIAL DISASTER DECLARATION FLOOD CONGRUENCE

	Baseline (1)	Pop. congruence (2)	County congruence (3)
Panel A. Event study coefficients			
Year of flood: flooded	0.079***	0.064***	0.062***
	(0.012)	(0.012)	(0.015)
Year of flood: media neighbor	0.031***	0.024***	0.022***
	(0.006)	(0.007)	(0.006)
Population TV media market congruence		0.028*	
		(0.016)	
County TV media market congruence			0.041**
			(0.020)
R^2	0.206	0.206	0.206
Communities	268,996	268,996	268,996
Panel B. Implied insurance takeup			
Median congruence (conditional on media market flood)		0.36	0.28
Implied flooded effect at the median		0.087	0.073
Implied media neighbor effect at the median		0.034	0.033

Notes: Panel A, columns 1–3, display select coefficients from estimation of equation (2) with media neighbors on the 1980–2007 panel. Column 1 reproduces the specification of Figure 8, panel B. Columns 2 and 3 run the same specification, except add the population and county congruence variables. The population (county) congruence variable measures the proportion of the population (counties) in the media market hit by a PDD flood. Panel B calculates the total hit and media neighbor implied insurance take-up effects by summing the congruence effect at the median with the relevant year of flood effect.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Results

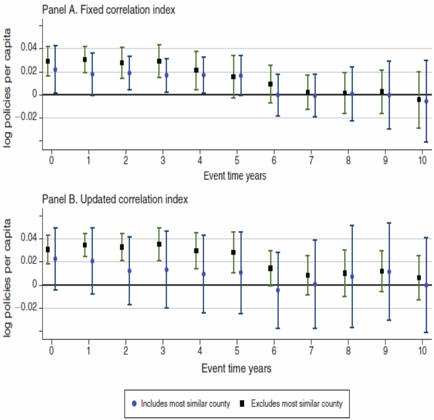


Figure 9. Insurance Take-Up for Media Neighbors by whether the Flood Includes the County with the Most Similar Flood History

Notes: Each panel plots media market neighbor insurance take-up coefficients from a distinct regression using a version of equation (2). The regression specification in the figure is the same as that of Figure 8, panel B, except that media market floods are divided into two types. The dots (squares) indicate insurance take-up for floods that include (exclude) the county with the most similar flood history. Panel A determines the county with the most similar flood history for each neighbor county using a flood correlation index and the years 1958–2007. Panel B defines the historical flood correlation index using only those years before the year of the current flood. Please refer to Section IIIB for further details.

Learning Interpretation

- Change and beliefs could be rational updating about flood risk
- Standard Bayesian learning model: Beta-Bernoulli
- p probability of flood in a community in a given year
  - Stationary distribution, independently drawn each year
- Distribution of p is $Beta\left(\alpha, \beta\right)$
- Homeowner conditional expectation of yearly flood probability p:

$$\mathbb{E}\left[p|S_t, t\right] = \frac{S_t + \alpha}{t + \alpha + \beta}$$

- t number of yearly observations
- $S_t = \sum_{s=1}^{t} y_s$ number of floods
- α, β parameters defining initial beliefs over flooding
- Implication: as stock of information t grows, effect of new observation goes to zero.

Learning Interpretation

- Event study graphs suggest homeowners are not considering all past flood information.
- Possilibities
  - Homeowners don't observe past info
  - Homeowners forget
  - [Another: homeowners incorrectly believe floods are serially correlated or flood risks are changing]
- Approach: weighting parameter that discounts past information.
  - Then conditional expectation of future flood can jump

Learning Interpretation

- Learning model with discounting:

$$\mathbb{E}\left[p|S_t', t'\right] = \frac{S_t' + \alpha}{t' + \alpha + \beta}$$

- $S_t^{'} = \sum_{s=1}^{t} y_s \delta^{t-s}$ weighted flood observations
- $t' = \sum_{s=1}^{t} \delta^{t-s}$ yearly observation equivalents. $\delta \in [0, 1.05]$
- $\delta < 1$ implies older flood matter less than recent floods when updating beliefs
- When $\delta = 1$, this reduces to full information model

Learning Interpretation

- Parameters α and β determine initial beliefs over flood probability.
- How determine? Three approaches
  - Mean county in country
  - Mean county in state
    - These match first two moments of empirical flood distribution with first two moments of Beta distribution
  - Flood distribution of county, allow certainty to vary

Learning Interpretation

- Simulate conditional flood probabilities using panel of PDD floods with Full Info and Discounting models

$$\mathbb{E}\left[p|S'_t, t'\right] = \frac{S'_t + \alpha}{t' + \alpha + \beta}$$

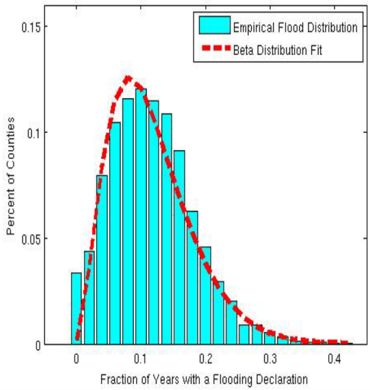
  - Approach: use this equation to simulate probability time series given different weighting values δ
  - Then choose time series of flood probabilities $p\left(\delta\right)_{ct}$ that minimizes mean square error of

$$\ln\left(takeup_{ct}\right) = \alpha + \beta_t \ln p\left(\delta\right)_{ct} + \alpha_c + \gamma_{st} + \varepsilon_{ct}$$

  - Same as baseline equation but replace event time variables with log flood probability

Learning Interpretation

Figure 4: Distribution of US Counties by Likelihood of a Presidential Disaster Declaration Flood from 1958-2007



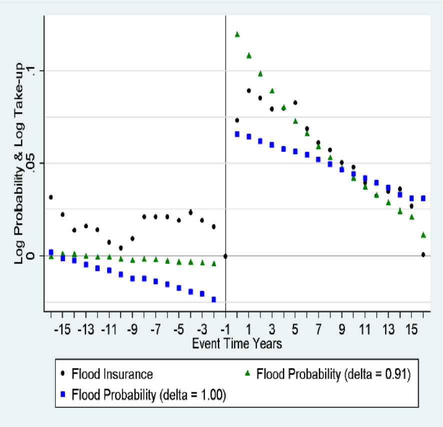
The figure graphs the yearly probability of a Presidential Disaster Declaration (PDD) Flood in each US county for the 50 years between 1958-2007. The histogram includes data from all 2704 counties in the 1980-2007 panel. The horizontal axis measures the fraction of years with a PDD flood and is calculated for each county by dividing the number of years with at least one PDD flood by 50. Each vertical bar shows the percent of counties with each 50 year flooding probability. The dashed curve is for a Beta Distribution with parameter values alpha = 2.87 and beta = 21.87. The parameter values are selected by matching the first two moments of the Beta Distribution to the first two moments of the empirical flood distribution shown in the histogram.

Learning Results

- Full Information Bayesian model can't match insurance takeup
  - Can't match speed of jump and decline to baseline
  - Jump too small, decline too small
- Best fitting model always has discounting ($\delta < 1$)

Learning Results

Figure 5: Flood Insurance Take-up and Simulated Probabilities



This figure plots the event study coefficients from three separate regressions using manuscript Equation (1) and the 1980-2007 panel. The three regressions differ only by the dependent variable: log insurance take-up (circles), and log simulated probability from the Full Information (squares) and Discounted (triangles) Bayesian model. The conditional probabilities are from a model that sets the initial values (prior beliefs) equal to the true expectation using historical flooding in the county and a value for $\alpha = 6.2$. Refer to Appendix Section F and Section 5 of the manuscript for more details.

Learning Results

- Migration alternative
  - Only use flood information if reside in county at time of flood
  - Calibrated using IRS county migration data.
  - Mean 1980-2007 migration rate is 5.5 percent
- Again best-fitting model has $\delta < 1$
- But $\delta = 1$ can't be rejected statistically

Learning Results

- Interpretation
  - Availability Bias (non-Bayesian interpretation)
  - Homeowners could mistakenly believe that past floods are less relevant for current flood risk
  - Homeowners could be learning and forgetting
  - Learning about the past may be costly

Conclusions

- This paper takes a widely-known fact: many people buy insurance right after calamity strikes
  - Simple question: what model of behavior can rationalize this?
  - Standard learning model can't accomodate it without a "forgetting" parameter
- Note: relevance for climate change.

Introduction: Lipscomb and Mobarak

- How does decentralization interact with externalities?
  - Crime
  - Traffic
  - Pollution

Setting: Water Pollution

- Over a billion people in the world lack sufficient potable water
- 1.5 million kids die each year from diarrheal disease
  - 17 percent of under-5 mortality
- 80 percent of sewage and 70 percent of industrial waste is dumped into surface waters untreated
- Hundreds of international and intranational conflicts over water use and management
- Coase Theorem: conflicts should be resolved through negotiation
  - Paper's goal: show that decentralization leads to negative spillovers, even within states

Setting: Water Pollution

- Data: 5,989 water quality measures, 372 upstream-downstream pairs
- All of Brazil's 8 river basins
- Overlay pollution data on GIS maps of moving county boundaries
  - Number of counties grew from 3,991 in 1980 to 5,507 in 2000
  - Use this to construct research design
  - Concerns with OLS: geographic and demographic differences between small v. large jurisdictions; endogenous placement of monitors; political differences between counties

Setting: Water Pollution

- Four predictions from model
  - Pollution increases towards downstream border
  - Pollution increases at increasing rate as we go downstream
  - Inflection or structural break in slope of pollution function at border
  - Pollution along a river increases with more border crossings

Relevance

- Decentralization en vogue for health, education, other public services in poor countries
  - World Bank, UNDP, etc.

Context

- Brazil: 15% of world's freshwater
  - 18 percent of counties have open sewers
  - 10 percent have industrial dumping
- Counties can influence environmental policy
  - Fine/tax pollution
  - Forbid some activities
  - Zoning policies (e.g., housing near rivers without sewers)
  - Environmental licenses

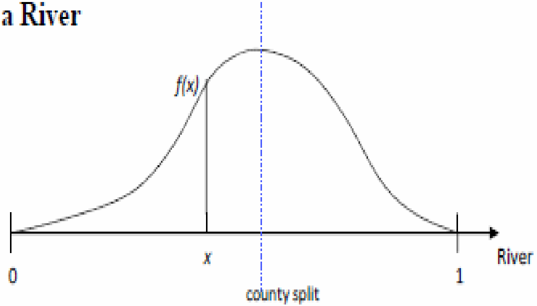
Context

- Creating new counties
  - A county has several districts, which can split off
  - Requires feasibility study, referendum, and state law
  - County transfer from federal government give incentive to create new counties
- Why do splits occur?
  - Disagreements over spending
  - Political differences within countiess

Model

- River flows left to right, population pdf $f(x)$
- Each person at location x consumes q_x, gets utility $u(q_x)$.
- Consuming q_x generates q_x units pollution

Figure 1: Model of a River



Model

- Pollution at x affects people downstream, with decay function at point t equal to
$$q_x e^{-(t-x)}$$
- q_x is both pollution and consumption
- Planner chooses consumption (=pollution) to tradeoff local utility against downstream pollution.
- Constraint: pollution never exceeds natural limit $\overline{q}$

Model

- Case 1: before jurisdictional split. Planner's problem: maximize aggregate utility from consuming q_x minus downstream disutility

$$\max_{q_x} f(x)\, q(q_x) - \int_x^1 q_x e^{-(t-x)} f(t)\, dt \qquad \text{s.t. } q_x \leq \overline{q}$$

- Downstream matters because in Case 1, it's part of the planner's jurisdiction
- Implied FOC, where λ is Lagrange multiplier on $\overline{q}$ constraint

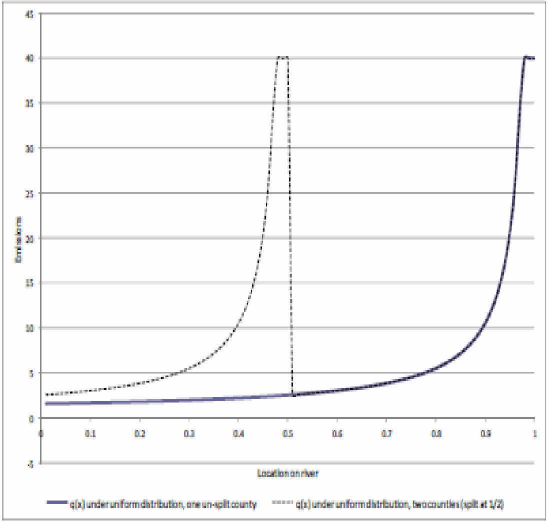
$$f(x)\, u'(q_x) = \int_x^1 e^{-(t-x)} f(t)\, dt + \lambda$$

- Analytical solution for uniformly distributed population of mass 1 and log utility $[u(q_x) = \ln(q_x)]$. Optimal pollution at x is

$$q_x^* = \min\left(\frac{1}{1 - e^{-(1-x)}}, \overline{q}\right)$$

Model

Figure 2: Emissions (q_x) Allowances Before and After the County Split

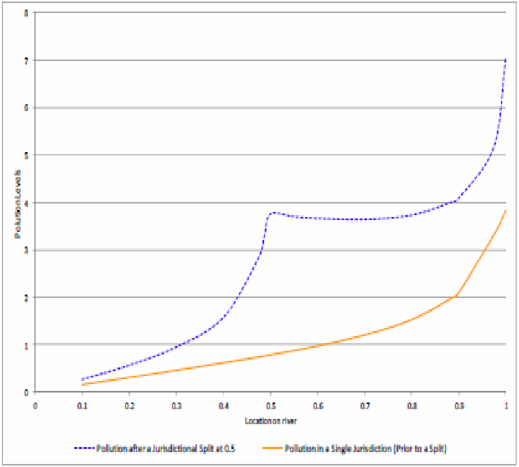


Model

- Actual pollution at a point y includes decayed upstream pollution:
$$P(y) = \int_0^y q_x^* e^{-(y-x)} f(x)\, dx$$

Model

Figure 3: The Pollution Function $P(y)$ for a County with Uniformly Distributed Population before and after the Jurisdictional Split at 0.5

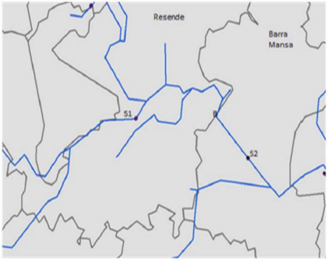


Model

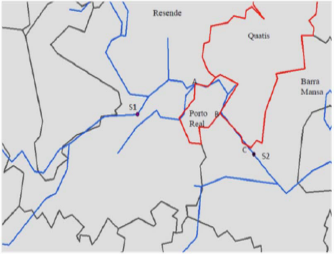
- Testable predictions in sum
  - Pollution increases as river travels downstream (positive slope)
  - Pollution increases at increasing rate
  - Inflection in pollution at border
  - Pollution increases more if a river crosses more borders

Identification Strategy

Figure 4: Example of the Evolution of County Boundaries in the State of Rio de Janeiro



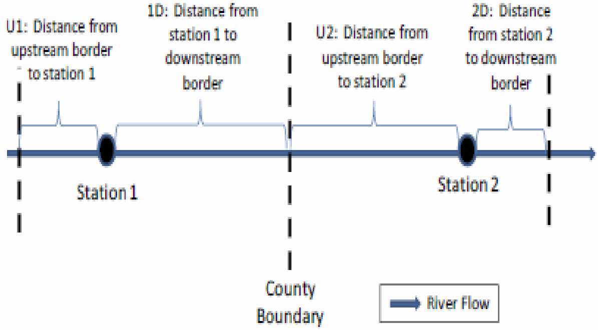
1991 Map, IBGE



1998 Map, IBGE

Identification Strategy

Figure 5: Notation on Distances



Data

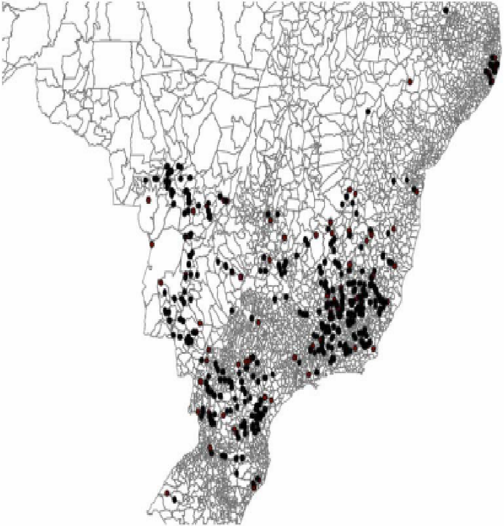
- 8,878 observations on biochemical oxygen demand (BOD)
- Reflects pollution from industry, sewage, and runoff

Data

Figure 6: Rivers and Water Quality Monitoring Stations

Data

Figure 7: Water Quality Monitoring Stations and County Boundaries



Data

Table 1: Summary Statistics

	Obs	Mean	Std. Dev.
Border Variables			
Number of border crossings	5989	8.422	11.218
Distance of station 1 from its nearest upstream border	5989	4.701	8.289
Distance from station 1 to its downstream border	5989	5.759	7.666
Distance of Station 2 from its nearest upstream border	5989	9.634	13.856
Distance from Station 2 to its downstream border	5989	12.358	18.731
County Characteristics			
GDP in Downstream County in 100,000 Reais	5989	4.388	17.343
Population in Downstream County in 100,000 people	5989	0.648	2.023
Area in Downstream County in 100,000 square kilometers	5989	0.025	0.040
GDP in upstream county in 100,000 Reis	5989	3.897	16.629
Population in Upstream County in 100,000 people	5989	0.559	1.721
Area in Upstream county in 100,000 square kilometers	5989	0.020	0.022
Water Quality			
BOD at upstream station	5989	3.529	12.529
BOD at downstream station	5989	3.382	15.247

- Context: clean water has BOD$<$2; polluted water has BOD of 4 to 8; treated sewage has BOD$=$20, raw sewage has BOD$=$300

Econometric Model

Station-level:

$$
\begin{aligned}
& \ln BOD_{i,t} \\
= \quad & \alpha_i + \sum \delta_{basiN-month} + \sum \gamma_{basiN-year} \\
& + \beta_1 Dist\_Upstr\_Border\_to\_Station(U1)_{i,1} + \beta_2 U1^2_{i,t} \\
& + \beta_3 Dist\_Station\_to\_Downstr\_Border(1D)_{i,t} + \beta_4 1D^2_{i,t} \\
& + \sum_k \lambda_k X^k_{i,t} + \alpha_i \theta_t + \varepsilon_{it}
\end{aligned}
$$

- Pollution variation depends on station FE; seasonal and annual variation; time-varying distance to nearest upstream (U1) and downstream(U2) border
- Distance variables of interest: $\beta_1, \beta_2, \beta_3, \beta_4$

Station fixed effects means these are only identified from changes in border locatoins over time

Econometric Model

Station-pair level:

$$
\begin{aligned}
& \ln\left(BOD^2_{i,t}\right) - \ln\left(BOD^1_{i,t}\right) \\
= \quad & \alpha_i + \alpha_i\theta_t + \sum \delta_{basiN-month} + \sum \gamma_{basiN-year} + \beta_1 Border\_Crossng \\
& +\beta_2 Dist\_Station1\_to\_Downstr\_Border(1D)_{i,t} \\
& +\beta_3 Dist\_Upstr\_Border\_to\_Station2(U2)_{i,t} \\
& +\beta_4 U1_{i,t} + \beta_5 2D_{i,t} + \sum_k \lambda_k X^k_{i,t} + \varepsilon_{i,t}
\end{aligned}
$$

Results

Table 2: Station Level Regression: Determinants of Changes in Pollution at a Station

Dependent Variable: 100*LogBOD in the upstream station

Distance from station 1 to Downstream border (1D)	-1.996***	-2.007***	-1.654**	-2.016**	-1.978**	-2.902**	-3.141**
	(0.729)	(0.758)	(0.684)	(0.853)	(0.766)	(1.234)	(1.267)
Squared Distance from station 1 to Downstream border ($1D^2$)	0.033***	0.033***	0.029***	0.033***	0.032***	0.034*	0.037*
	(0.011)	(0.011)	(0.009)	(0.011)	(0.011)	(0.019)	(0.020)
Distance from Upstream border to station 1 (U1)	-0.250	-3.316	-3.618	-3.337	-3.416	0.983	-5.024
	(0.926)	(3.163)	(2.784)	(3.228)	(3.166)	(1.946)	(5.251)
Squared Distance from upstream border to station 1 ($U1^2$)		0.186	0.181	0.152	0.190		0.409
		(0.191)	(0.180)	(0.222)	(0.191)		(0.376)
Distance follows border downstream of station 1					-0.111		
					(0.266)		
Distance follows border upstream of station 1					-0.111		
					(0.530)		
GDP of the county in which the station is located in 100000 Reis	0.354	0.341	0.249	0.221	0.341	-0.007	-0.013
	(0.345)	(0.347)	(0.341)	(0.346)	(0.347)	(0.669)	(0.661)
Population of the county, in 100,000	2.270	1.687	1.068	3.448	1.670	5.155	3.394
	(2.000)	(2.101)	(2.236)	(2.526)	(2.106)	(6.422)	(6.691)
Area of the county, in 100,000 square km	-322.743**	-274.982**	-236.013**	-259.219**	-278.598**	-319.828**	-264.123**
	(124.828)	(114.928)	(114.129)	(118.044)	(117.466)	(94.830)	(120.403)
Basin*year fixed effects	Y	Y	N	N	Y	Y	Y
Basin*month fixed effects	Y	Y	N	N	Y	Y	Y
Month* year fixed effects	N	N	Y	N	N	N	N
Basin*year*month fixed effects	N	N	N	Y	N	N	N
Station trends	N	N	N	N	N	Y	Y
Observations	5,989	5,989	5,989	5,989	5,989	5,989	5,989
R-squared	0.065	0.065	0.060	0.098	0.065	0.134	0.135
Number of pair	372	372	372	372	372	372	372
F-stat for slope of pollution function upstream = slope downstream (evaluated at 1km from border)	3.159	2.415	3.071	2.209	2.427	0.735	2.046
Prob > F	0.076	0.121	0.081	0.138	0.120	0.392	0.153
F-stat for slope of pollution function upstream = slope downstream (evaluated at 3km from border)	2.812	2.841	4.219	3.007	2.833	0.017	2.232
Prob > F	0.094	0.093	0.041	0.084	0.093	0.896	0.136
F-stat for slope of pollution function upstream = slope downstream (evaluated at 5km from border)	2.574	3.163	5.970	3.649	3.125	0.038	2.084
Prob > F	0.109	0.0761	0.015	0.057	0.078	0.845	0.150

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The dependent variable is 100*the log level BOD at the upstream station. All regressions include station pair fixed effects. Standard errors are clustered by the station. All regressions include an indicator variable for stations which are not separated by a border, and the distance 1D and U2 (the distance from station 1 to the nearest downstream border and the distance from the nearest upstream border to station 2 are set equal to 0). An indicator variable is also included for cases in which there were no intermediate counties for GDP variables (cases of 0 or 1 crossings).

Results

Table 3: Station Pair Regressions: Change in Pollution from Upstream to Downstream Station Over time

Dependent Variable: Log Difference in BOD (Dowstream minus Upstream)*100

Number of borders crossed between station 1 and station 2	3.200*	3.595**	3.418**	3.886**
	(1.647)	(1.536)	(1.494)	(1.550)
Distance from station 1 to Downstream border (1D)		1.526*	1.771*	1.740*
		(0.885)	(1.029)	(1.028)
Distance from Upstream border to station 2 (U2)		-3.101	-3.347	-3.442
		(1.965)	(2.108)	(2.097)
Outside Station Pair Control Variable: Distance from Upstream border to station 1 (U1)		-1.676	-1.597	-2.179
		(2.244)	(2.244)	(2.257)
Outside Station Pair Control Variable: Distance from station 2 to its Downstream border (2D)		-0.115	0.119	0.076
		(0.522)	(0.606)	(0.609)
Distance river follows the border to upstream station			0.401	0.590*
			(0.331)	(0.344)
Distance river follows the border to downstream station			1.472	1.302
			(2.274)	(2.273)
Distance river follows the border between stations			0.403	0.320
			(0.640)	(0.624)
Distance river follows the border downstream of downstream station				-0.000
				(0.001)
Distance river follows the border upstream of upstream station				-0.001
				(0.001)
GDP of the county in which the upstream station is located in 100000 Reis	0.027	-0.233	-0.224	-0.252
	(0.693)	(0.650)	(0.674)	(0.678)
GDP of the county in which the downstream station is located in 100000 Reis	-0.048	0.009	0.001	0.006
	(0.195)	(0.206)	(0.205)	(0.209)
Population of the upstream county, in 100,000	-6.456	-4.722	-4.024	-3.973
	(7.469)	(7.162)	(7.492)	(7.509)
Population of the downstream county, in 100,000	19.427***	24.111**	20.912**	21.507**
	(5.399)	(9.619)	(9.979)	(10.276)
Area of the upstream county, in 100,000 square km	136.379	234.667	-8.623	7.237
	(210.705)	(204.781)	(438.153)	(430.683)
Area of the downstream county, in 100,000 square km	-144.256**	82.426	114.246	145.269
	(68.011)	(195.500)	(227.955)	(230.920)
Average GDP in intermediate counties in 100,000 R$ in constant 2000 R$	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Average Population in intermediate Counties in 100,000 people	-0.000	-0.009	-0.011	-0.012
	(0.009)	(0.014)	(0.014)	(0.014)
Average Area in 100,000 square kilometers	-0.230***	-0.190***	-0.207***	-0.180**
	(0.049)	(0.047)	(0.067)	(0.077)
River basin*year FE	Y	Y	Y	Y
River basin*month FE	Y	Y	Y	Y
Station pair Trends	Y	Y	Y	Y
Observations	5,989	5,989	5,989	5,989
R-squared	0.120	0.122	0.122	0.122
Number of pair	372	372	372	372
F-test for slope of pollution function upstream of border = slope downstream of border		3.150	3.014	3.122
Prob > F		0.0773	0.0840	0.0786

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The dependent variable is 100*the log difference in BOD between the downstream station and the upstream station. All regressions include station pair fixed effects, station pair trends, river basin-year, and river basin month dummies. All regressions also include controls for GDP, population, and area in the upstream county, the downstream county, and the average in the intermediate counties. Standard errors are clustered by the downstream station. All regressions include an indicator variable for stations which are not separated by a border, and the distance 1D and U2 (the distance from station 1 to the nearest downstream border and the distance from the nearest upstream border to station 2 are set equal to 0. An indicator variable is also included for cases in which there were no intermediate counties for GDP variables (cases of 0 or 1 crossings).

Results

Table 4: Non-linearity in the Pollution Function, at Varying Distances from Borders

	Split at x=1	Split at x=3	Split at x=5	Split at x=15
U2 - far from border (Distance from upstream border to station 2 * beyond *x* km of station)	-2.523	-3.129	-2.624	-1.290
	(1.990)	(1.957)	(1.987)	(1.460)
U2 - close to border (Distance from upstream border to station 2 * within *x* km of station)	-9.510	-12.887***	-4.347	-4.040
	(18.927)	(4.621)	(4.262)	(3.509)
1D - close to border (Distance from station 1 to downstream border within *x* km of station)	9.372	7.765	0.446	-0.734
	(15.566)	(4.722)	(2.235)	(2.118)
1D - far from border (Distance from station 1 to downstream border * beyond *x* km from station)	1.941*	2.125**	1.794*	1.609*
	(1.066)	(1.000)	(1.075)	(0.956)
2D Distance from station 2 to downstream border	-0.025	0.124	-0.009	-0.916
	(0.485)	(0.501)	(0.494)	(0.799)
U1 Distance from upstream border to station 1	-2.452	-2.369	-2.453	-2.158
	(2.278)	(2.197)	(2.310)	(2.275)
Observations	5,989	5,989	5,989	5,989
R-squared	0.120	0.121	0.120	0.120
Number of pair	372	372	372	372
F-stat for slope upstream of border = slope downstream of border (U2 close = ID close)	0.628	8.418	0.808	0.558
Prob > F	0.429	0.00410	0.370	0.456
F-stat for test of non-linearity in slope (1D close = 1D far)	0.230	1.442	0.565	1.865
Prob > F	0.632	0.231	0.453	0.173

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The dependent variable is 100*the log difference in BOD between the downstream station and the upstream station. All regressions include station pair fixed effects, station pair trends, river basin-year, and river basin month dummies. GDP, population, area controlled in all specifications but not shownStandard errors are clustered by the downstream station. All regressions include an indicator variable for stations which are not separated by a border, and the distance 1D and U2 (the distance from station 1 to the nearest downstream border and the distance from the nearest upstream border to station 2 are set equal to 0). An indicator variable is also included for cases in which there were no intermediate counties for GDP variables (cases of 0 or 1 crossings).

Results

Table 5: Station Pair Regressions with controls for post-split Counties

Dependent Variable: Log Difference in BOD (Dowstream minus Upstream)*100

Number of borders crossed between station 1 and station 2	4.785**	5.258***	4.966***	5.494***
	(1.994)	(1.576)	(1.507)	(1.519)
Distance from station 1 to Downstream border (1D)		1.022	1.386	1.368
		(0.927)	(1.103)	(1.096)
Distance from Upstream border to station 2 (U2)		-4.630*	-4.769*	-4.847**
		(2.527)	(2.467)	(2.447)
Outside Station Pair Control Variable: Distance from Upstream border to station 1 (U1)		-1.899	-1.791	-2.536
		(3.367)	(3.347)	(3.201)
Outside Station Pair Control Variable: Distance from station 2 to its Downstream border (2D)		-0.083	0.209	0.160
		(0.500)	(0.597)	(0.589)
Distance river follows the border to upstream station			-0.055	0.286
			(0.471)	(0.539)
Distance river follows the border to downstream station			2.332	2.104
			(2.180)	(2.195)
Distance river follows the border between stations			0.490	0.473
			(0.688)	(0.683)
Distance river follows the border downstream of downstream station				0.000
				(0.001)
Distance river follows the border upstream of upstream station				-0.002
				(0.001)
New county post split, downstream	-6.120	-31.682*	-27.819*	-26.062*
	(14.428)	(16.645)	(14.235)	(14.425)
Old county post split downstream	9.614	2.385	0.115	1.620
	(16.903)	(18.613)	(19.316)	(19.684)
New county post split upstream	10.324	12.414	9.835	9.765
	(19.023)	(22.660)	(21.250)	(21.770)
Old county post split upstream	24.975*	22.617	22.450	22.552
	(15.013)	(16.106)	(16.934)	(17.080)
GDP of the county in which the upstream station is located in 100000 Reis	0.381	0.313	0.339	0.322
	(0.750)	(0.805)	(0.825)	(0.835)
GDP of the county in which the downstream station is located in 100000 Reis	-0.086	-0.093	-0.104	-0.104
	(0.182)	(0.176)	(0.176)	(0.181)
Population of the upstream county, in 100,000	-5.028	-7.791	-6.358	-5.609
	(8.072)	(9.177)	(9.524)	(9.631)
Population of the downstream county, in 100,000	17.566*	10.954	8.212	8.813
	(9.244)	(10.734)	(11.525)	(11.562)
Area of the upstream county, in 100,000 square km	-96.025	-41.334	-487.624	-547.910
	(195.025)	(176.710)	(489.395)	(493.358)
Area of the downstream county, in 100,000 square km	-253.363*	130.387	177.779	222.210
	(144.184)	(270.482)	(291.354)	(300.591)
Average GDP in intermediate counties in 100,000 R$ in constant 2000 R$	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Average Population in intermediate Counties in 100,000 people	0.007	-0.002	-0.003	-0.005
	(0.011)	(0.015)	(0.016)	(0.015)
Average Area in 100,000 square kilometers	-0.323***	-0.241	-0.215	-0.111
	(0.124)	(0.162)	(0.188)	(0.210)
River basin*year FE	Y	Y	Y	Y
River basin*month FE	Y	Y	Y	Y
Station pair Trends	Y	Y	Y	Y
Observations	5,989	5,989	5,989	5,989
R-squared	0.120	0.122	0.122	0.122
Number of pair	372	372	372	372

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The dependent variable is 100*the log difference in BOD between the downstream station and the upstream station. All regressions include station pair fixed effects, station pair trends, river basin-year, and river basin month dummies. All regressions also include controls for GDP, population, and area in the upstream county, the downstream county, and the average in the intermediate counties. Standard errors are clustered by the downstream station. All regressions include an indicator variable for stations which are not separated by a border, and the distance 1D and U2 (the distance from station 1 to the nearest downstream border and the distance from the nearest upstream border to station 2 are set equal to 0. An indicator variable is also included for cases in which there were no intermediate counties for GDP variables (cases of 0 or 1 crossings).

Mechanisms etc.

Table 6: Robustness Check: Station-Level Regression

	Dependent Variable: 100*LogBOD in the upstream station			
	Top 1% dropped		Top 1% top coded	
Distance from station 1 to Downstream border (1D)	-2.323***	-3.297***	-2.056***	-3.130**
	(0.848)	(1.247)	(0.776)	(1.254)
Squared Distance from station 1 to Downstream border ($1D^2$)	0.035***	0.039**	0.033***	0.037*
	(0.012)	(0.020)	(0.011)	(0.019)
Distance from upstream border to station 1 (U1)	-4.276	-4.926	-3.659	-5.059
	(2.997)	(5.181)	(3.073)	(5.241)
Squared Distance from upstream border to station 1 ($U1^2$)	0.238	0.411	0.204	0.411
	(0.185)	(0.372)	(0.188)	(0.375)
GDP of the county in which the upstream station is located in 100000 Reis	0.677***	0.220	0.409	0.002
	(0.184)	(0.192)	(0.318)	(0.551)
Population of the upstream county, in 100,000	-3.638	1.263	0.312	3.197
	(2.387)	(4.592)	(2.048)	(6.020)
Area of the upstream county, in 100,000 square km	-206.116*	-226.390*	-252.729**	-251.029**
	(107.581)	(128.719)	(110.863)	(122.854)
Includes station pair trends	N	Y	N	Y
Observations	5,940	5,940	5,989	5,989
R-squared	0.070	0.137	0.067	0.135
Number of pairs	372	372	372	372

The dependent variable is 100*the log level BOD at the upstream station. All regressions include station pair fixed effects. Standard errors are clustered by the station. All regressions include an indicator variable for stations which are not separated by a border, and the distance 1D and U2 (the distance from station 1 to the nearest downstream border and the distance from the nearest upstream border to station 2 are set equal to 0). An indicator variable is also included for cases in which there were no intermediate counties for GDP variables (cases of 0 or 1 crossings).

Mechanisms etc.

Table 7: Robustness Check: Station-Pair Regression

	Dependent Variable: 100* (Log downstream BOD-Log Upstream BOD)			
	Top coded 1% of observations		Top 1% of observations dropped	
Number of borders crossed between station 1 and station 2	3.522**	3.639**	2.949**	3.063**
	(1.518)	(1.502)	(1.328)	(1.315)
Distance from station 1 to Downstream border (1D)	1.452	1.444*	1.594*	1.591**
	(0.898)	(0.863)	(0.820)	(0.785)
Distance from Upstream border to station 2 (U2)	-3.204*	-3.005	-2.908*	-2.710*
	(1.786)	(1.853)	(1.511)	(1.543)
Outside Station Pair Control Variable: Distance from Upstream border to station 1 (U1)		-1.723		-1.639
		(2.209)		(2.061)
Outside Station Pair Control Variable: Distance from station 2 to its Downstream border (2D)		-0.127		-0.150
		(0.512)		(0.497)
Observations	5,989	5,989	5,913	5,913
R-squared	0.117	0.117	0.114	0.114
Number of pairs	372	372	372	372
F-test for slope of pollution function upstream of border = slope downstream of border	3.792	3.245	4.931	4.267
Prob > F	0.0528	0.0730	0.0274	0.0400

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The dependent variable is 100*the log difference in BOD between the downstream station and the upstream station. All regressions include station pair fixed effects, station pair trends, river basin-year, and river basin month dummies. All regressions also include controls for GDP, population, and area in the upstream county, the downstream county, and the average in the intermediate counties. Standard errors are clustered by the downstream station. All regressions include an indicator variable for stations which are not separated by a border, and the distance 1D and U2 (the distance from station 1 to the nearest downstream border and the distance from the nearest upstream border to station 2 are set equal to 0. An indicator variable is also included for cases in cases in which there were no intermediate counties for GDP variables (cases of 0 or 1 crossings). In the first three regressions BOD observations for both the upstream and downstream stations above the 1 percentile observation are coded at that observation (41.7).

Mechanisms etc.

Table 8: Causes of and Responses to Water Pollution

Panel A. County-Reported Causes of Water Pollution	Count*
Mining	235
Oil and gas from boats	81
Animal Waste	832
Materials from the Processing of Sugar	160
Industrial Dumping	521
Domestic Sewage	1595
Poor Solid Waste Management	821
Poor enforcement of river pollution regulations	648
Poor enforcement of underground water rights licensing	228
Use of Pesticides and Fertilizers	901
Others	160
Total Counties reporting Water Pollution	2121
Panel B. County Actions to Reduce Pollution	
Fining Households with Inadequate Sewer Systems	2462
Fining Companies with Inadequate Industrial Waste Management System	1007
Monitoring of Potentially Polluting Industrial Activities	596
Taxing Mining Industries	1027
Taxing Automobiles	104
Management of Toxic Waste	483
Trash Collection Program	1654
Recycling Program	1082
Creation of Sewers	1949
Other	564

*Counts are as of 2002. There were 5,560 counties in Brazil in 2002. Source: IBGE Environmental Census

Mechanisms etc.

Table 9: Non-point Source Pollution Measures (Placebo Check)

	Log Total Dissolved Top 1% dropped		Log Turbidity Top 1% dropped		Log Conductivity Top 1% dropped	
Distance from station 1 to Downstream border (1D)	-0.833	0.163	-0.756	-0.246	-0.604	-0.358
	(1.725)	(1.801)	(1.400)	(1.861)	(0.928)	(1.509)
Squared Distance from station 1 to Downstream border ($1D^2$)	-0.027	-0.018	0.074***	0.055**	-0.019	-0.031
	(0.024)	(0.024)	(0.025)	(0.028)	(0.016)	(0.026)
Distance from Upstream border to station 1 (U1)	-2.486	-1.827	0.545	1.094	2.468	2.849
	(3.436)	(3.615)	(5.035)	(6.689)	(2.696)	(3.265)
Squared Distance from upstream border to station 1 ($U1^2$)	0.395*	0.284	0.036	-0.356	-0.326*	-0.164
	(0.235)	(0.238)	(0.278)	(0.424)	(0.192)	(0.179)
GDP of the county in which the upstream station is located in 100000 Reis	0.146	0.121	0.363	1.007***	0.262	-0.041
	(0.185)	(0.207)	(0.634)	(0.228)	(0.300)	(0.201)
Population of the upstream county, in 100,000	1.740	-0.017	-1.479	-14.056**	-1.452	-4.173
	(2.722)	(3.699)	(7.015)	(5.844)	(4.053)	(6.149)
Area of the upstream county, in 100,000 square km	50.663	-145.122	572.230	828.424**	-262.508	-118.067
	(194.715)	(387.520)	(427.286)	(321.383)	(293.611)	(372.832)
Includes station pair trends	N	Y	N	Y	N	Y
Observations	4,882	4,882	5,725	5,725	5,651	5,651
R-squared	0.212	0.276	0.367	0.421	0.133	0.228
Number of pairs	300	300	363	363	363	363

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

All regressions include station pair fixed effects. Standard errors are clustered by the station. All regressions include an indicator variable for stations which are not separated by a border, and the distance 1D and U2 (the distance from station 1 to the nearest downstream border and the distance from the nearest upstream border to station 2 are set equal to 0). An indicator variable is also included for cases in which there were no intermediate counties for GDP variables (cases of 0 or 1 crossings).