Price Incentives for Conservation: Experimental Evidence from Groundwater Irrigation

Nick Hagerty
Ariel Zucker *

August 15, 2025

Abstract

Policymakers often seek to reduce resource consumption but face constraints that preclude the use of prices or other first-best instruments. This paper studies the design and effectiveness of an alternative approach: payments for voluntary conservation. We offered payments for reduced groundwater use among farmers in Gujarat, India in a randomized controlled trial. Price incentives work: The program reduced irrigation time by 22 percent. Conservation payments are a practical policy tool in this setting: They saved energy at a per-unit cost comparable to the electric utility's supply costs. Contract design greatly affects cost-effectiveness: More stringent benchmark values (against which conservation is measured and rewarded) halved the average cost of conservation.

*Hagerty: Montana State University. Zucker: UC Santa Cruz. We are grateful to Jasmin Korat, Mustufa Patel, Gokul Sampath, Neha Doshi, and Getsie Immanuel for outstanding project management and research assistance; Piyush Gandhi and Brooke Banning for excellent research assistance; Gaddala Ganesh, Aditya Madhusudan, Isha Fuletra, and Suresh Bharda for excellent project assistance and field support; and many survey enumerators with J-PAL South Asia for data collection. We thank Naveen Patidar and the staff of the Aga Khan Rural Support Programme (India) and the Aga Khan Foundation for implementation support. We are also grateful for feedback from countless colleagues, including workshop and seminar participants at MIT, IPWSD, UCSB, the World Bank, AERE, UCSC, Heartland, EPIC, WEAI, AAEA, and PERC. Funding for this research was generously provided by the King Climate Action Initiative, the International Growth Centre, the Weiss Fund for Research in Development Economics, the Abdul Latif Jameel World Water and Food Security Lab (J-WAFS) at MIT, and the PERC Lone Mountain Fellowship. This research has IRB approval from the Institute for Financial Management and Research and is registered on the AEA RCT registry (ID number AEARCTR-0009337).

1 Introduction

Across a wide range of settings, natural resources are overused, degrading the quality and quantity of the resources themselves and harming others through negative externalities. Policymakers trying to solve these problems often seek to reduce resource use but face constraints that rule out textbook "first-best" policy instruments. Simple pricing or corrective taxes may be politically untenable (Pai and Strack, 2022), while creating new property rights or markets may be prohibitively costly (Libecap, 2014). Common alternatives, such as rationing or subsidies for resource-saving technologies, tend to be inefficient or expensive. There is a need for alternative "second-best" tools that can achieve conservation goals while operating within real-world political and administrative constraints.

One alternative gaining wider use is payments for voluntary conservation, in which a government or other principal offers payments to users who reduce their resource consumption (Wolak, 2006; Jayachandran et al., 2017; Aspelund and Russo, 2025). Such a program can replicate the marginal incentives of a Pigouvian tax, at least for some users, so it may achieve similar outcomes. But because it relies on rewards rather than penalties, it reverses the distributional consequences, potentially relaxing policy constraints. How well do conservation payment programs work? How cost-effective are they? And how can they best be designed?

This paper studies the design and effectiveness of a conservation payment program. We experimentally evaluate a program of payments for groundwater conservation among farmers in Gujarat, India – a setting where marginal incentives to conserve are minimal, yet the consequences of overuse are severe. Groundwater is an essential resource for irrigation and drinking water worldwide, but unregulated extraction in many regions has led to depletion, increasing poverty and conflict while reducing farm income, wealth, and employment (Sekhri, 2014; Blakeslee et al., 2020). The problem is compounded by electricity pricing: power used to pump groundwater is often free, flat-rate, or heavily subsidized, exacerbating depletion while straining public utilities and degrading electricity service (Burgess et al., 2020). Yet high energy subsidies are often seen as valuable means of redistribution; in India, reform efforts are often met with political resistance and public protest (Sovacool, 2017).

Our program installed meters and offered payments for reduced groundwater pumping during the main irrigation season of 2022-23 within a randomized controlled trial. The basic design is to (1) meter the groundwater pumps of all study participants, (2) offer randomly selected participants payments for reduced pumping relative to a "benchmark" quantity, and (3) compare groundwater consumption by these farmers to that of the rest of the sample. The program was implemented in collaboration with the Aga Khan Rural Support Programme (India), a trusted organization with a long history in the study area.

We first ask how this program affected groundwater and electricity consumption. We find that payments for groundwater conservation work. Farmers offered conservation payments irrigated for 22 percent fewer hours than farmers assigned to the control group, with a 95 percent confidence interval of $(12, 33)$ percent. The effects are similar when we convert irrigation time to energy use—we estimate that treatment farmers reduced electricity consumption by 140 kilowatt-hours (kWh) per month relative to the control-group mean of 611 kWh. Treatment effects increased over the three months of the intervention, suggesting a durable response. Higher prices (relative to lower prices) reduced electricity use but not pumping duration, suggesting that farmers with more powerful pumps are more price-sensitive.

Then, we assess the cost-effectiveness of payments for groundwater conservation from the perspective of an electric utility – the natural choice to implement a similar program at scale. At the margin, is it cheaper for a utility to increase supply or to reduce demand through conservation payments? We find that our program is cost-competitive with typical costs of electricity procurement in northwestern India. Our program spent 6.1 INR in total conservation payments for every kWh of energy saved, a cost that is slightly greater than the average costs of electricity provision for the utility in our study area (and may be lower than marginal costs) and slightly lower than costs in a nearby state. Considering the additional social costs of groundwater depletion and of emissions from electricity generation, it appears likely that paying farmers to reduce their groundwater irrigation would bring greater social benefits than purchasing more electricity and distributing it for free.

Finally, we study the optimal design of contracts in our setting. In particular, we ask how to choose benchmarks – the values against which conservation is measured and rewarded. Benchmarks can greatly influence program cost-effectiveness, because they determine payment expenditures and can affect conservation behavior. Often a natural choice of benchmark is a measure of historical resource use, since it may be a good predictor of resource use absent the program. But a natural choice is not necessarily a cost-effective choice. We show theoretically that cost-effective benchmarks are generally different from either historical use or the predicted counterfactual. We then show empirically in our setting that more stringent benchmarks greatly improve cost-effectiveness.

More stringent benchmarks are cost-effective for two reasons. First, they reduce inframarginal payments – rewards for resource units that never would have been consumed anyway. We show in our setting that inframarginal payments are large but would have been even larger had we used historical consumption as benchmarks (instead of targeting stringency on expected variance). Second, more stringent benchmarks can actually induce more conservation. In our treatment group, we also cross-randomized participants into high- and low-benchmark groups, which separated benchmarks by about 30 percent. Many participants responded to lower benchmarks by conserving more, yielding a seemingly "free lunch" of both lower costs and greater benefits. Overall, the average cost of conservation in the low-benchmark group was less than half the cost in the high-benchmark group.

One central contribution of this paper is to provide experimental evidence that marginal prices can reduce groundwater irrigation. The basic idea is core to microeconomic theory, but empirical evidence has been limited because there have been so few real-world examples to study. We build most directly upon two non-randomized pilots of similar programs run by electric utilities in India; Fishman et al. (2016) found no effects in Gujarat, while Mitra et al. (2023) found water use reductions in Punjab. By working instead with a local non-governmental organization, we are able to randomize participants, directly measure pumping at the individual level, and trace the demand curve to prices beyond what utilities have been willing to test. In the most closely related randomized study, Chakravorty et al. (2023) induce a volumetric price using an encouragement design; they do not find effects on water use, which may be because relatively few farmers adopted the price.[1]

Observational research on irrigation prices has faced two key challenges: finding exogenous variation in groundwater prices, and reliably measuring water use. Early studies in the U.S. used proxies for pumping costs and relied on self-reported water use (Gonzalez-Alvarez et al., 2006; Hendricks and Peterson, 2012; Pfeiffer and Lin, 2014). A few more recent papers study the introduction of explicit prices. One in West Bengal uses self-reports of water use and finds inconclusive results (Meenakshi et al., 2013). Three in the U.S. are able to observe meter readings from individual wells and find reductions in water use (Smith et al., 2017; Bruno and Jessoe, 2021; Bruno et al., 2024). Introducing randomized variation allows us to more cleanly isolate causal effects, while installing meters allows us to directly observe irrigation behavior in a setting beyond the western U.S.

Our project also contributes to the broader literature on payments for environmental services (PES). Our intervention has the same basic structure as hundreds of programs designed to incentivize the provision of environmental services, ranging from increased forest or wetland cover, to reduced input intensity in agriculture.[2] Despite their preva-

[1] They also show that a water-saving technology saves water only in regions with marginal prices; although this technology is randomized, the heterogeneity is not.

[2] For example, in the United States alone, payments are available to farmers for actions to mitigate flood

lence, rigorous evaluation of these types of programs has been limited (see Pattanayak et al. (2010) and Börner et al. (2017) for reviews). Most existing evaluations use covariate matching and are unable to address selection bias, a particular concern for a voluntary program. The exceptions are four randomized controlled trials of programs to improve land management in Colombia (Pagiola et al., 2016), reduce deforestation in Uganda (Jayachandran et al., 2017), encourage tree planting in Malawi (Jack and Cardona Santos, 2017), and reduce crop burning in India (Jack et al., 2023). Our study shows that PES models are feasible and can be effective in a novel context: reducing energy and water use in agriculture.

Finally, we contribute to literature connecting the price response of electricity consumption in developing countries to policy decisions about energy-sector investment and reform. Experimental and quasi-experimental studies are still limited, but a few have been conducted recently on rural households in Columbia (McRae, 2015), urban households in South Africa (Jack and Smith, 2016), and new grid connections in Kenya (Lee et al., 2018).

2 A Conceptual Model of Conservation Payments

We study a program to incentivize resource conservation using the following contract. A principal offers a positive payment to an agent that depends on their resource consumption. The contract specifies three parameters, a price p, a benchmark b, and a maximum payment $\bar{y}$. It makes a positive payment y that pays p for every unit of resource forgone, measured relative to the benchmark:

$$y = \begin{cases} 0 & q \geq b \\ p \cdot (b - q) & b > q > b - \frac{\bar{y}}{p} \\ \bar{y} & q \leq b - \frac{\bar{y}}{p} \end{cases}$$

If consumption exceeds the benchmark, the payment is zero. If consumption is less than the benchmark, the payment equals the price times the difference between quantity consumed and the benchmark. If the maximum payment is reached, further conservation does not increase the payment. Figure 1 illustrates the budget set of this contract.

and wildfire risks, provide habitat for endangered species, salinity mitigation, and water and energy conservation.

2.1 How do conservation payments compare with a per-unit tax?

The conservation payments contract imposes the same marginal incentives, at least over a certain range of consumption. But its structure also differs in two important ways. First, the contract effectively includes a lump-sum transfer; for agents who consume less than the benchmark (and below the maximum payment), the contract is equivalent to a linear tax $-pq$ plus a transfer pb. Second, this tax applies only within a certain range. Agents who consume more than the benchmark, or less than the maximum payment, face no marginal incentives. Two kinks in the budget set ensure the payment is nonnegative and below the maximum but complicate the agents' response.

2.2 How do agents respond to conservation payments?

Not all agents change their behavior in response to conservation payments. Because the contract provides a linear incentive only for quantities below the benchmark, agents who would have consumed above the benchmark without the program may continue to consume the same quantity. This differs from a per-unit tax, in which all agents are marginal to the incentive, in the sense that any positive quantity consumed is subject to a per-unit price. Under conservation payments, some agents are marginal, but not all.

To see this, Figure 1 plots quasi-linear indifference curves over quantity consumed (including both private benefits and costs) and conservation payments for two example agents. Without conservation payments, the budget set is flat and coincides with the x-axis; with conservation payments, the budget set is piecewise linear. Agent A changes her resource consumption in response to the conservation payments program. Her indifference curves are tangent to the x-axis at q_0^A and tangent to the conservation payments budget set at q_1^A, so her resource consumption is lower under conservation payments relative to the counterfactual. Agent B does not change his resource consumption when offered conservation payments; his indifference curves are tangent to both budget sets at q_B.

2.3 How does the benchmark affect conservation?

The choice of benchmark can affect resource consumption through two possibly opposing channels: the extensive margin and the intensive margin.

First, the extensive margin: More lenient (i.e., higher) benchmarks induce more agents to respond to the program and conserve the resource at all. If the benchmark is sufficiently high, all agents respond – everyone optimizes on the sloped segment of the budget constraint and receive positive payments. If the benchmark is sufficiently low, no agents respond – the payments are too far away to be worth pursuing. The program is unattractive if agents must complete too much uncompensated conservation before payments kick in, or if the available range of compensated conservation (above zero or the maximum payment) is too small.

In Figure 1, relaxing the benchmark (i.e., raising it) would be represented by a parallel shift outward in the budget constraint. For any well-behaved set of indifference curves, the maximum utility reachable on the sloped segment is greater after it is shifted outward. Any agent already on the sloped segment will continue to optimize on the sloped segment, and some agents on the flat segment will move to the sloped segment.

Second, the intensive margin: Conditional on conserving at all, more lenient benchmarks may lead agents to *increase* consumption – to conserve *less* than they otherwise would. On the intensive margin, relaxing the benchmark is a pure income effect, so if the resource is a normal good, the agent increases both resource consumption and the payment.[3] A similar effect may also operate through behavioral channels. More stringent (i.e., lower) benchmarks may increase conservation if agents put more weight on the cash rewards than on the costs of forgone resource consumption, or if they interpret benchmarks as containing information about scientific recommendations or social norms.[4]

Because there are plausible effects in opposite directions, the net effect of benchmarks on conservation is theoretically ambiguous. Appendix A presents a slightly more formal analytical model with proofs of these effects for two broad classes of utility functions.

2.4 How does the benchmark affect program cost-effectiveness?

We can define the cost-effectiveness of the program to the principal as the average expenditure on conservation payments per unit of conservation achieved by the program. Here, conservation is measured relative to the counterfactual without the program ("causal conservation"), not relative to the benchmark ("rewarded conservation").

Benchmarks affect cost-effectiveness in two ways. First, benchmarks mechanically affect payment expenditures. More lenient benchmarks require rewarding more units of conservation, regardless of whether those units would have been consumed anyway. Benchmarks set higher than counterfactual consumption require inframarginal (or "non-additional") payments. Benchmarks set lower than counterfactual consumption may

[3]Of course, an increase in consumption is not the only possibility. If utility is quasilinear, the benchmark does not affect the quantity because there are no income effects, and if the resource is an inferior good, raising the benchmark leads the agent to consume less of it.

[4]Our program made no such claims, but participants still could make these interpretations.

achieve some units of conservation for free. This effect favors more stringent benchmarks.

Second, benchmarks affect the conservation achieved by the program, as described above. Putting all effects together, more lenient benchmarks induce more agents to conserve but increase inframarginal payments. More stringent benchmarks lead some agents to "give up" conserving, but they reduce payment expenditures and may induce greater conservation through income and/or behavioral effects.

3 Study Setting and Experimental Design

3.1 Enrollment and Sample

We implemented a randomized controlled trial among groundwater-irrigating farmers in Saurashtra, a water-scarce region of Gujarat state, India. The study villages are located in the inland districts of Rajkot, Surendranagar, and Morbi (shown in Figure 2). Groundwater depletion is a concern within the study area, and nearby areas are marked by some of the most rapid groundwater depletion rates both within India and globally (Jasechko et al., 2024). While the primary source of employment in the study area is in agriculture (Registrar General and Census Commissioner of India (2001)), there are also a number of industrial occupation opportunities.

We recruited our sample using lists of villagers currently or formerly participating in agricultural outreach programs with our implementing partner, the Aga Khan Rural Support Program (AKRSP), and its sister agency the Aga Khan Foundation (AKF). The outreach programs included Better Cotton Initiative, which aims to improve the sustainability of the global cotton supply; Farmer Producer Groups, which aim to empower farmers in marketing produce and procuring high-quality inputs; and various micro-irrigation subsidy and support programs. Surveyors approached farmers on these lists, as well as any farmers who shared water with those on the lists, to determine eligibility.

In order to be eligible for the study, the household's primary agricultural decision-maker (PAD) was required to meet the following criteria: Planted crops and irrigated primarily using groundwater in the previous Rabi season; planned to irrigate during the coming Rabi season; had no more than two active wells on their primary farm; had electric-powered pumps on all active wells; did not have multiple pump starters in use on any active well; and did not belong to a network of sharing irrigation sources among groups of farmers larger than four.

We enrolled a total of 1,347 farmers who met the eligibility criteria, completed a baseline survey, and consented to the full study (including installation of an hours of use meter on the pumpsets used to irrigate their primary farm). Of these, 236 attrited prior to randomization, and another 122 prior to the final data collection visit, leaving an analysis sample of 989 farmers.

3.2 Hours-of-Use Meters

For all participants, an hours-of-use meter was installed on the electric pump starter of their primary irrigation source or sources.[5] The meter measures the cumulative time, to tenths of an hour, that the pump has been in operation. The difference in values displayed on the meter at two different points in time allows us to measure a farmer's total irrigation time during that period. Meters were installed in Fall of 2022, and were read monthly by survey staff from December 2022 through March 2023 (Figure 4).

Participants were able to remove the meters – we found through piloting that this was key to broad acceptability.[6] However, removing a meter would reduce data quality and could allow a participant to obtain higher conservation payments by reducing the recorded irrigation time. We therefore implemented several practices to both detect and discourage removal. We ensured easy detection by attaching custom stickers to the easiest disconnection points such that disconnection would tear the sticker. If meter removal or tampering was detected at any meter-reading visit, participants were immediately disqualified from receiving further payments. We explicitly requested participants to keep the meter installed through the end of the irrigation season, and we rewarded participants with 100 INR per meter for keeping their meters installed without tampering through the final meter reading.

Throughout the enrollment and meter-installation process, we informed all participants that some of them would be offered a new program that involved cash payments in exchange for water conservation, and that participants in this new program would be chosen by lottery. Participants were informed of their treatment assignment at the end of the first meter reading visit in December 2022, so we were able to collect one month of irrigation data before participants learned of their treatment status.

[5] Farmers in our sample had up to two wells on their primary farm, and therefore up to two metered pump starters.

[6] We used hours-of-use meters because in piloting activities they performed better and were more acceptable to farmers than water or electricity meters. Water meters were expensive, prone to clogging with debris, required custom hardware for each farmer due to non-standardized irrigation equipment, reduced irrigation flexibility, often raised water pressure and therefore pumping costs, and were easy to disconnect. Electricity meters raised suspicion that our team was working with the local electric utility; hours-of-use meters are functionally nearly identical but avoided such suspicion. Throughout enrollment and meter installation, we characterized our project as a research study about how to encourage water conservation, and we emphasized that the individual readings would not be shared with any utility or government agency.

3.3 Experimental design

The experiment had two overarching treatment arms: *conservation payment* farmers were eligible to receive payments for conserving groundwater below a benchmark, whereas *control* farmers received no such incentives. Figure 3 illustrates our experimental design.

Treatment: Conservation payments Participants in the treatment ("Conservation Credits") group were offered payments for conserving water during the following three months of the winter growing season, known as the rabi season: January to March. Rabi is the peak irrigation season in the region; as there is typically no rainfall, agriculture is entirely dependent on irrigation. At each meter reading, participants were informed of their benchmark for the following month, and the payment for the previous month was calculated. Payments were awarded at a fixed rate for consuming fewer hours of irrigation than the monthly benchmark, according the formula:

$$\text{Payment}_{it} = \min\left(\max\left(0, \text{price}_i \times ((\text{hours benchmark})_{it} - (\text{hours consumed})_{it})\right), (\text{max payment})_i\right) \quad (1)$$

where price_i is the per-hour incentive rate, $(\text{hours benchmark})_{it}$ is an individual-month-specific benchmark, $(\text{hours consumed})_{it}$ is the monthly meter reading, and $(\text{max payment})_i$ is the maximum monthly payment.[7] Payments were pro-rated in the case that meter readings were not exactly 31 days apart. The payments were later disbursed via electronic bank transfer.

Sub-treatment arms Within the overall treatment group, we randomly assigned participants to one of four sub-treatment arms. These arms differ along two dimensions: the per-hour incentive rate and the benchmark. Individuals assigned a *high price* received 100 INR (1.20 USD) per hour conserved, and those assigned a *low price* received 50 INR (0.60 USD) per hour conserved. The prices were chosen to encompass realistic ranges of groundwater prices that a policymaker might wish to set. The low price represents the approximate cost of electricity provision for the median farmer and is similar to the price offered in a program in Punjab (Mitra et al., 2023).[8] The high price allows us to study the

[7]The maximum monthly payment was 4,000 INR for farmers with one well and 6,000 INR for farmers with two wells. These maximums were not pro-rated.

[8]A price of 50 INR per hour is approximately equal to the unsubsidized average cost of electricity supply in Gujarat for the median pumpset in our sample. That is: (5.4 INR/kWh average cost of electricity provision in Gujarat) * (5 hp pump brake power) / (40% typical motor efficiency) * (0.75 kW/HP conversion factor) = 50 INR/hr. The Punjab program offered an incentive of 4.0 INR/kWh, which translates to different per-hour prices for different farmers depending on pump power, but would be approximately 37 INR per hour for the median pumpset in our sample. For more details on this calculation see Section 4.2

response to prices well beyond those piloted by electric utilities in India to date, which might be justified by the additional social costs of groundwater depletion and electricity generation.

Individualized benchmarks were set using a formula that sought to optimize the expected number of marginal farmers as a function of first-month pumping data (i.e., after meter installation but before treatment assignment was revealed to surveyors or farmers). Individuals assigned the *high* and *low benchmark* received 115% and 85%, respectively, of their formula-based benchmark, rounded to the nearest 10-hour increment.

Control Participants in the Control group were informed that they were not selected for the new payment program but still had their meters read monthly for four months and were offered the reward for keeping the meters installed for the duration of the project.

3.4 Randomization

Randomization was conducted at the level of farmer-sharing group: that is, the set of farmers who mentioned at the baseline survey that they used any common irrigation sources. By randomizing at the sharing group level, we minimize the possibility that conservation payments will spillover to control farmers.

Randomization was stratified by forecasted hours of irrigation and size of water-sharing group. Specifically, the final sample of water-sharing groups was ordered first by number of farmers, and second by forecasted hours of irrigation.[9] Groups were randomly allocated in equal proportion between the control and treatment arms using a pseudo-random number generator (Stata software) within each ordered pair. Pairs were then combined into ordered cells of eight farmer-sharing groups, within which the four groups allocated to the treatment arm were randomly allocated in equal proportion between the four sub-treatment arms.

4 Data and Summary Statistics

4.1 Data Sources

Our analysis rests on data from two primary sources: a baseline survey and meter reading data. First, we conducted a baseline survey with both self-reported and field mea-

and Table 4.

[9] Forecasted hours of irrigation were created using a random forest using baseline survey data and geological mapping data. Forecasts were fit using a sample of farmers in Saurashtra from a previous project.

surement components prior to randomizing participants into treatments. Self-reported data include demographic and socioeconomic characteristics, such as landholding size and household size; cropping, crop management, and irrigation decisions in the previous year; the power of the primary pumpsets; and water conservation strategies and attitudes. Field measurements include the precise geolocation and depth-to-water of each well on the participant's largest farm where measurement is safe and feasible. We also collect the names and contact details of any farmers who use water from the primary farm or whose water is used on the primary farm in order to sort our sample into water-sharing groups consisting of all farmers who are connected through water sharing relationships. Baseline data was collected electronically through tablet surveys.

Second, we directly measure groundwater pumping for all study participants using hours-of-use meters installed on the pump starter of each participant's primary irrigation source.[10] Surveyors recorded meter readings each month using a digital tablet survey. Meter data quality was assured through random audits, in which a research associate compared the digitally recorded meter readings with dated, geo-located photographs of the meter dial included on the tablet survey.

We combine these with two secondary data sources: hydrogeology data from digitized "Groundwater Prospects Maps" created by the Government of India and proxies of agricultural yields derived from satellite data. Specifically, we first supplement our baseline data with an extensive set of hydro-geological features from the "Groundwater Prospects Maps" prepared by the National Remote Sensing Center, Government of India. These maps are available only as images; we digitize the map images and convert them to vector data that can be used in analysis. Following Ryan and Sudarshan (2022), we extract variables on rock type, aquifer type, and fractures that are predictive of groundwater availability and yields in hard rock areas aquifer systems.

We supplement both our baseline and outcome data using agricultural yield proxies derived from satellite imagery. We utilize two common proxies for agricultural biomass: the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI). These indices are calculated using Google Earth Engine from Sentinel-2 images at a 10 meter resolution (which are taken every five days) in a 40 meter radius around the coordinates of participants' wells. We filter out pixels of clouds and surface water, and use standard techniques to correct for data anomalies from atmospheric disturbances.

[10] Analog hours-of-use meters manufactured by Nishant Engineers (model: NE53/6S).

4.2 Outcome Variables

Our primary outcome variable is monthly hours of groundwater irrigation. Meters show cumulative duration of pump operation, so we calculate monthly irrigation hours as the difference between values shown on the meter at the current visit and previous month's visit. Because not all meter-reading visits occurred at exact monthly intervals, we rescale observed hours to a 31-day rate so that observations are comparable across farmers and months.

Our secondary outcome variable is energy consumption in irrigation. Energy use is not observed directly but rather converted from hours of irrigation using known functional relationships from physics. The formula is:

$$E = \frac{P_b}{\eta_m} \times t \tag{2}$$

where E is energy consumed, t is duration of pump operation, P_b is the power rating of the pump's motor ("brake horsepower"), and η_m is the motor efficiency, a unitless constant between zero and one.[11]

We collect t and P_b in meter-reading and baseline surveys. Motor efficiency η_m is difficult to measure accurately and would have required use of electricity meters, with which many study participants were uncomfortable. Instead, we draw an estimate of motor efficiency from the literature in the most similar setting we can find: 40 percent (Mitra et al., 2023).

Note that our energy use variable is not simply a monotone transformation of irrigation hours, since it also depends on pump power, which varies across farmers. That said, pump power does not endogenously respond to the program,[12] so energy use can be seen as a rescaling of units combined with a reweighting of farmers within the sample. Either way, average treatment effects may be substantively different if individual treatment effects are heterogeneous and correlated with pump power. Similarly, individual-level price elasticities of demand for energy and hours are equal, but aggregate price elasticities may differ due to this reweighting.

A third set of outcomes measures the economic impacts of water conservation using two proxies for agricultural yields: log seasonal differences of EVI and NDVI for each

[11] To obtain E in kilowatt-hours (kWh) when P_b is measured in horsepower (hp), the formula also requires the unit conversion of 0.7457 kW per hp.

[12] In a longer-term incentive program, farmers could make investments in new pumps. Our program lasted for only one irrigation season, and participants were unable to replace a pump without removing the meter and becoming disqualified from the program.

farmer. For each index, we follow Asher and Novosad (2020) and proxy productivity over the 2022-23 rabi season (December to March) as the maximum index value reached on the farmers' land during the season less mean index value in the first four weeks of the season.[13] Our preferred outcome is the natural log of this difference, which eases interpretation. We calculate rabi productivity using the same methodology in the year prior to the intervention (2021-22) and mean index values in November 2022, prior to randomization, as additional baseline controls.

4.3 Descriptive Statistics and Balance Checks

Descriptive statistics. Table 1 reports baseline characteristics of the experimental sample. In both this table and all subsequent analysis, the sample is restricted to farmers who completed all rounds of data collection: the baseline survey, meter installation, the baseline meter reading, and all three meter-reading visits during the intervention.

Our sample consists predominantly of smallholder farmers; the mean plot area of their primary farm is 1.95 hectares.[14] Most participants are literate, have completed primary and secondary education, and identify with a "scheduled caste/scheduled tribe/other backward caste" designation. Only half own a plow or tractor. Cotton is the primary crop in our sample, with sorghum/millet, groundnut, and pulses as the next most common crops. Farmers are at least somewhat diversified in their crops, with a mean count of distinct crops around 2.

Most participants have only one active well; some have two. Some wells are dugwells and others are borewells (tubewells); the most common type of well is a dug-*cum*-borewell, in which a borehole is drilled into the bottom or side of a dugwell in order to access additional pockets of water. The average well is 59 meters deep, but many are considerably deeper. The most common electric pump installed in each well is rated at 5 horsepower, but some are more powerful.

Many farmers in our sample are already using cultivation practices that conserve water. 41 percent use a drip irrigation system, 69 percent use raised beds, and 19 percent use rotational, strip, or zero tillage. Local water markets are rare in our context: Only 1 percent report having purchased water for irrigation. Farmers sometimes share irrigation sources with neighbors, usually relatives, but water sharing is not a large share of

[13] Daily index values are first averaged across wells for farmers with multiple wells. We difference the mean index value of the first four weeks of the growing season, as opposed to six weeks in Asher and Novosad (2020), to avoid contamination by the start of our intervention.

[14] The Indian government typically defines farmers holding less than 2 hectares as "small and marginal farmers".

irrigation in our sample: About 10 percent of pump operation during the previous (2021-22) irrigation season was directed to irrigation off the primary farm, which includes both neighbors and secondary farms also held by the respondent.

Balance. Columns 3 and 4 of Table 1 report means of baseline characteristics separately for the overall treatment and control groups. The two groups appear similar across all characteristics. We formally check for balance test between the main treatment and control groups using a Wald F-test for joint orthogonality of all characteristics reported in this table. The F-statistic is small and the p-value is large, so we fail to reject the null hypothesis that treatment-control differences are zero for all characteristics.

Our sample includes more farmers in the treatment group than in the control group, implying that attrition rates were different across the two groups. Differential attrition would bias the results if attrition is correlated with characteristics that predict the outcome variable. But we do not see evidence that the treatment and control groups are differentially selected across a range of baseline characteristics.

5 Program evaluation

We first evaluate our conservation payments program as implemented by estimating intent-to-treat (ITT) effects of eligibility for the overall intervention. We estimate post-double-selection LASSO regressions of the following form:

$$Y_{it} = \alpha_t + \tau \cdot \text{ConservationCredits}_i + \gamma' \mathbf{X}_{it} + \varepsilon_{it}, \tag{3}$$

where Y_{it} is a outcome variable for farmer i at monthly visit t, and $\text{ConservationCredits}_i$ is an indicator for being in the overall treatment group and therefore eligible for payments. X_{it} is a vector of individual-specific covariates interacted with month t chosen by double-selection LASSO (Belloni et al., 2013); Appendix B lists the full set of candidate covariates.[15] Standard errors are clustered by randomization pair (following de Chaisemartin and Ramirez-Cuellar (2024)), which nests months within farmer and farmers within groups of neighbors that reported sharing water prior to the intervention.[16]

[15] We implement post double-selection LASSO in Stata with the commands dsregress for OLS, dspoisson for Poisson, and the user-written command ivlasso (Ahrens et al., 2018) for instrumental variables regressions, respectively.

[16] We omit randomization pair fixed effects following Bai et al. (2023), who show that they complicate the interpretation of the estimand and do not necessarily reduce bias from differential attrition.

Because our primary outcome variables are right-skewed, we expect covariates to have multiplicative rather than additive effects on the outcomes. We therefore estimate the same specifications using Poisson pseudo-maximum likelihood, which can more precisely estimate regression-adjusted treatment effects in such cases (Chen and Roth, 2024):

$$Y_{it} = \exp\{\alpha_t + \tau \cdot \text{ConservationCredits}_i + \gamma' \mathbf{X}_{it}\} \cdot u_{it}, \tag{4}$$

Poisson regression directly recovers a transformation of the average treatment effect as a proportion of the control mean (Silva and Tenreyro, 2006).[17]

5.1 Conservation payments reduce irrigation time and energy use

Table 2 presents estimated effects of overall payment eligibility on hours of irrigation, the variable directly measured by our meters. Our preferred estimate is the covariate-adjusted specification in column (3): Farmers assigned to the program operated their pump for an average of 10.4 fewer hours per month during the intervention period than control-group farmers. This effect represents a 22 percent reduction relative to the control-group mean of 47 hours per month, and the 99% confidence interval excludes zero.

Results are broadly robust to alternative specifications. Column (7) shows the same specification as column (3) but estimated using Poisson regression, again indicating that the program led to a 22 percent reduction ($e^{-0.25} - 1$) in irrigation hours. Estimates are less precise without covariate adjustment (columns (1) and (5)) but the 90% confidence intervals still exclude zero. Columns (2) and (5) include village fixed effects and no other covariates, a parsimonious example of a fully saturated regression model. We include this specification to confirm that the main result is similar in a specification guaranteed to be unbiased for the average treatment effect even in finite samples (Athey and Imbens, 2017).

To see how the program affected the distribution of irrigation hours, Figure 5 plots bin treatment effects – i.e., effects of payment eligibility on the share of participants whose irrigation hours fall within certain ranges – using a subset of controls. These estimates are less precise but important patterns are visible. Figure 5(a) plots bin effects by hours of irrigation. It shows that the program moved participants from higher values of pumping to lower values. Figure 5(b) plots bin effects by hours relative to the benchmark, which

[17] Our primary outcome variables sometimes take a value of zero, so we cannot run log-linear OLS regressions. We avoid other "log-like" transformations, such as the inverse hyperbolic sine, because they are sensitive to the choice of units (Mullahy and Norton, 2023) and because any notion of an individual-level percentage change is undefined for a variable that admits zero (Chen and Roth, 2024).

varies across participants. It shows that the program moved participants from above their benchmark to below it, suggesting that participants understood the program structure and complied with its incentives.[18]

Effects on energy use are shown in Table 3. Again our preferred estimate is in column (3): Payment eligibility reduced energy use by 140 kWh per month, a 23 percent reduction compared with a control-group mean of 611 kWh per month. Alternative specifications are broadly consistent though less precise than the irrigation hours regressions. Point estimates without covariate adjustment are much smaller but so imprecise that we cannot reject equality with our preferred specification.

5.2 Treatment effects increased over time

To investigate seasonal patterns in treatment effects, we augment our primary regressions to estimate separate treatment effects in each month of the program. Results are plotted in Figure 6 for both OLS and Poisson estimates of our preferred specification, which includes controls selected by double LASSO.

Average effects of the program increased in magnitude over the course of the experiment, from 5 hours in the first month to 12 and 13 hours in the last two months of the program. (We can reject that the first and second, or the first and third months, are equal, at a 5 percent significance level.) These differences are even more dramatic when expressed as a percentage of the control mean, which declined over time. Treatment effects estimated using Poisson regression increased from 9 percent in the first month to 22 and 28 percent in the second and third months.

We see two likely explanations for the growing response over time. One possible reason is increasing trust in the program. Because the conservation payments program was a completely new concept, it seems likely that participants would have changed their behavior only tentatively in the first month. After they saw real cash appear in their bank accounts, they responded less cautiously. Another possible reason is that demand for irrigation becomes more elastic later in the growing season. For many crops, water application is most critical during an early phase of growth. After this early phase, yields may be less sensitive to irrigation amounts, and so farmers would become more sensitive to the price of irrigation. We do not currently have data to distinguish between these explanations, but we expect both are at play.

[18]Usually we would not expect to see bunching immediately below the benchmark threshold since the budget set is concave at this point; anti-bunching is more likely. In this case, the reason that effects concentrate in the first bin below the threshold is likely due to the zero lower bound. Most benchmarks are set in the 10-50 hour range, so most participants can never appear in the first three bins shown in the graph.

5.3 Higher prices affect energy use more than irrigation time

Next, we go beyond the effects of the program overall to investigate whether the level of price incentive affects irrigation behavior, conditional on program participation. We compare the high- and low-price sub-treatment groups by interacting the overall treatment variable with an indicator for being in the high-price subgroup. The results are in columns (4) and (8) of Tables 2 and 3.

Across specifications and outcomes, the main effects of the program on irrigation hours remain large and statistically significant, while the interaction effects are smaller and not statistically significant. This says that being offered a price incentive of 50 INR per hour, relative to not being offered a price incentive at all, has a greater effect on conservation than increasing the price from 50 to 100 INR per hour. This result is consistent with a convex demand curve: There may be many low-cost opportunities to conserve water and energy resources that are left on the table when marginal resource prices are zero but adopted when prices are positive, but once that low-hanging fruit is picked, resource conservation faces more rapidly rising opportunity costs.

But for the outcome of energy use, this pattern is reversed. The interaction effect is greater than the main effect, meaning that we obtained more energy conservation from raising the price (from low to high) than from offering the program at all (with only the low price). Energy use is simply a rescaling of irrigation hours, not an independently measured outcome, and the only component of the scaling factor that differs across farmers is pump power. So this result implies that farmers with more powerful pumps responded more to the higher price than those with less powerful pumps. Perhaps for these farmers, the opportunity cost of an hour of pumping is larger, and a higher hourly price may thus be needed to encourage conservation.

Appendix D estimates a price elasticity of demand, using treatment eligibility as an instrument for effective price. While the response to our program is specific to our design parameters, a demand elasticity is potentially more generalizable. However, we hesitate to put much weight on this result. The problem is that some participants may have responded to the incentives even without receiving a payment, violating the exclusion restriction. For example, they might have generally reduced their consumption without paying close attention to the benchmark, or perhaps they reduced their consumption early in the month before deciding the benchmark was too stringent to be worth pursuing.

5.4 Conservation payments do not harm agricultural productivity

To understand whether the intervention has any measurable economic impacts, we use Equation 3 to examine the effects of the intervention on proxies of agricultural yields. Results are shown in Appendix Table 11.

Despite reducing irrigation intensity, payment eligibility does not lead to any significant change in yields. Our preferred estimate is in column (3): the seasonal increase in NDVI among payment eligible farmers is about 7.6 percent *larger* than among control farmers, and although this estimate is noisy we can reject substantial decreases in productivity among treated farmers. The results are similar if we use alternative measures of agricultural productivity. Overall, we find no evidence that the reductions in irrigation had negative economic consequences for farmers.

6 Cost-effectiveness

Now, we consider the cost-effectiveness of our conservation payment program from the perspective of an electricity utility. For now we set aside the social costs of groundwater depletion and focus solely on energy. Suppose political constraints rule out straightforward volumetric prices for electricity. Might a utility company find it less costly to reduce demand via a conservation payments program than to increase supply by procuring additional electricity?

We calculate the cost of reducing electricity demand through this program as the ratio of total expenditures on conservation payments to total energy conserved. Note this ratio is not just a rescaling of our demand estimates, because it includes payments made to inframarginal farmers. For total energy conserved as a result of the program, we use the preferred OLS estimate from Table 2 because it is more precise than the Poisson estimate. Table 4 shows further details of this calculation, parameters used, and results.

We estimate that conservation payment program reduced electricity use at a cost of 6.1 INR per kWh conserved. This value appears to be similar to published estimates of the costs of electricity procurement. It is slightly greater than the average cost of electricity procurement per unit sold by the electric utility in our study area, 5.4 INR per kWh, but the marginal costs of electricity procurement are likely greater than than average costs. It is also lower than the cost of electricity procurement in the nearby state of Punjab.

In this calculation, we only consider expenditures on conservation payments and omit other program costs such as meter hardware and personnel and travel expenses for reading meters. We do so for two reasons. First, electric utilities obtain other benefits from metering their customers, so we prefer to consider the perspective of a utility that is already collecting this data. Second, metering costs in a permanent program would likely be lower than in our short-term intervention. It would likely be more cost-effective to install smart meters that can be read remotely, saving the labor and travel expenses, the fixed cost of which would then be amortized over a longer period.

It is also worth considering the social costs of groundwater depletion and of air pollution from electricity generation. A utility company may not include these costs in a cost-effectiveness calculation, but a government may want to consider them as motivation for subsidizing a conservation payments program. Estimating the negative externalities from groundwater extraction is beyond the scope of this study. But even relatively small estimates of these social costs would likely make it socially optimal for a utility to offer conservation payments before expanding electricity supply.

7 Contract design: Choosing benchmarks

Last, we consider the choice of benchmark values against which conservation is rewarded. Many conservation programs set the benchmark equal to a historical measure of the target outcome – for example, the number of trees on a property at program enrollment, or the amount of water used last year. In some settings, historical use might reliably predict the target outcome absent the program, such that conservation measured relative to this value indeed represents conservation relative to the counterfactual. But in other settings, it might not predict the counterfactual well. And if not, then basing the benchmark on historical use might not be the best approach.

Assuming a principal seeks to maximize the cost-effectiveness of a conservation payment program, the cost-effective benchmark generally does not coincide with historical use nor the best prediction of counterfactual use. We suggest that cost-effective benchmarks are likely to be stringent: The optimal contract will often reward conservation relative to a benchmark value set substantially lower than historical use. This is because more stringent benchmarks reduce expenditures on inframarginal payments and also may induce more conservation.

7.1 Stringent benchmarks reduce inframarginal payments

First, we show that inframarginal payments are large in our setting. Inframarginal payments arise from prediction error – when the benchmark is set above resource use without the program, and they increase mechanically with the benchmark. This fact suggests a strategy of targeting benchmarks on counterfactual uncertainty. Perhaps benchmarks ought to be set approximately equal to the counterfactual when it is predictable, and lower when it is uncertain.

This logic inspired the design of the benchmarks that we implemented in the treatment group. We expected (and found) that pumping variance increases with baseline pumping (i.e., in the first month after meter installation and before randomization); Appendix Figure 9 shows this pattern of heteroskedasticity. We therefore set benchmarks approximately equal to first-month pumping at low values but lower (more stringent) at higher values, as the predictive value of first-month pumping fell. Figure 7 plots our benchmarks as a function of first-month pumping, with the $y = x$ line for comparison. Benchmarks start low at low values of first-month pumping, rise roughly one-for-one for a range of values, but then flatten out and decline again.[19] Lenient benchmarks would have been very expensive for participants whose first-month pumping exceeds 200 or 300 hours, so we chose not to chase them; some ended up pumping near zero during the program even absent incentives.

Our benchmark design reduced total payments by 45% relative to setting benchmarks equal to first-month pumping. We can estimate inframarginal payments by calculating what payments would have been in the control group (where there is no behavioral response) under alternative benchmark schemes. Appendix Table 12 lists average inframarginal payments calculated under each of several alternative benchmark scenarios. Inframarginal payments vary widely based on how benchmarks are set. They are also a large share of total payments, implying that inframarginal payments are an important driver of cost-effectiveness.

7.2 Stringent benchmarks induce more conservation

Next, we turn to the behavioral response. How does pumping respond to the benchmarks? Recall that more lenient benchmarks may induce participants to conserve on the extensive margin, but more stringent benchmarks may induce participants to increase conservation.

To test which effect dominates, we analyze our randomized benchmark sub-treatment.

[19] The broad shape of this function came from optimizing a nonlinear function of first-month pumping to maximize the share of participants who would be marginal to the incentive (i.e., pump at or below the benchmark) subject to the project's budget constraint. This function was increased or decreased by 15 percent in the high-benchmark and low-benchmark randomized sub-treatment groups. Benchmarks were rounded to the nearest 10 hours, set separately for each well, and summed. The minimum benchmark was 10 hours. Remaining variation comes from a small number of errors in field operations, as benchmarks were calculated during the same visit at which first-month pumping was observed.

Benchmarks were set approximately 30 percent higher for participants in the high-benchmark group than in the low-benchmark group. The first stage of this sub-treatment is strong: Table 5, column 1 shows that it increased benchmarks by 9 hours on average with an F-statistic of 135.

In the overall sample, evidence is not strong that random marginal changes to benchmarks affect pumping. In Table 5, column 2 reports a regression of irrigation hours on overall treatment (i.e., payment eligibility) and its interaction with the high-benchmark sub-treatment. Relative to the low-benchmark group, the high-benchmark group pumped about 5 hours more per month, but the confidence interval extends from −1 to 10.

However, there are strong reasons to expect the effect of benchmarks to differ across participants. At low values of water use, the extensive-margin effect is likely to dominate – small changes in benchmarks are large relative to water use, while income effects are small – so we would expect more stringent benchmarks to decrease conservation. At high values of water use, the intensive-margin effect is likely to dominate – benchmarks are too low to affect the extensive margin for most participants, while income effects are greater – so we would expect more stringent benchmarks to increase conservation.

This is in fact what we find: Participants with high expected pumping respond to lower benchmarks by conserving more. Columns 3-4 of Table 5 report the same regression estimated in "low pumping" and "high pumping" subsamples, which split the full sample at 60 hours of first-month pumping, the threshold where our benchmark formula flattens out.[20] For participants with low first-month pumping, the benchmark has a small and statistically insignificant effect. But for participants with high first-month pumping, there is a sizable positive effect – lower benchmarks reduce pumping. Columns 5-6 report 2SLS versions of the regressions; at least within the range of benchmarks we set, lower benchmarks appear to reduce pumping roughly one-for-one.

This pattern holds when we expand the analysis. Figure 8 plots the effect of the high-benchmark sub-treatment by seven quantiles of first-month pumping. These estimates come from a Poisson regression so that the effects are comparable to each other in proportional terms. For participants with low first-month pumping, a higher benchmark decreases pumping by around 20 percent. For participants with high first-month pumping, a higher benchmark increases pumping by 20 to 30 percent. These results imply that to maximize water conservation, the program should offer higher benchmarks for participants with low expected pumping (so that conservation is worthwhile for more of them)

[20] This dimension of heterogeneity was not pre-specified, so these results should be treated as exploratory rather than confirmatory. They are not necessary for the overall result that the low-benchmark group is more cost-effective, because the difference is driven more by the inframarginal payments than the behavioral response.

and lower benchmarks for participants with high expected pumping (because they can be induced to conserve more that way).

7.3 Stringent benchmarks can improve cost-effectiveness

Finally, we put the inframarginal payments and behavioral response together and calculate the overall cost-effectiveness of the program under different scenarios. We do not attempt to globally optimize the benchmarks, which would require extrapolating far beyond our sample variation. Instead, we restrict attention to the local variation introduced by our randomized high- and low-benchmark sub-treatment groups.

Table 6 calculates cost-effectiveness in the full treatment group and each benchmark sub-treatment group. The low-benchmark group reduces pumping at a cost of 71 INR per hour, which is 39 percent less expensive than the full sample and 59 percent less expensive than the high-benchmark group. This difference is driven both by inframarginal payments and the behavioral response.

We can also locally optimize benchmarks within the range of our experiment. We choose either the high or low benchmarks to maximize conservation, using the quantile estimates from Figure 8. Within each quantile bin, we keep participants in the high-benchmark group if the effect of higher benchmarks on pumping is negative, keep participants in the low-benchmark group otherwise, and drop participants in the other group. Using this carefully selected sub-sample, we then recalculate overall program effects (i.e., relative to the full control group) and payment expenditures.

In Table 6, the conservation-maximizing group achieves more conservation than the full sample while spending only slightly more on conservation payments. As a result, the average cost of conservation is only 86 INR per hour in the conservation-maximizing group, as compared with 117 INR per hour in the full sample, producing a 27 percent increase in cost-effectiveness.

This simulation only maximizes conservation, not cost-effectiveness. As it turns out, the low-benchmark group is even more cost-effective than the conservation-maximizing group. Jointly optimizing both conservation and payment expenditures would likely produce a schedule of benchmarks that is even more cost-effective.

8 Conclusion

This study finds that moderately sized incentives for groundwater conservation lead farmers to reduce groundwater irrigation by approximately 20 percent. Impacts increase over time, indicating that the response to incentives can be sustained. And this is a short-term response: Our program lasted for only one irrigation season and was introduced after crops were already planted. In a longer-term program, the response would likely be even greater, since farmers would be able to substitute crops and adjust other inputs.

These findings suggest that conservation payments, a policy solution similar to existing "payments for environmental services" programs, are an effective tool for managing groundwater and energy resources in India. In many settings — and perhaps especially in the setting of agricultural groundwater extraction in India — Pigouvian taxes may be politically infeasible. By exchanging corrective taxes for subsidies, conservation credits overcome the political barriers to taxing the agricultural sector, while still introducing marginal incentives for conservation. Thus, conservation credits may be a particularly promising policy approach for reducing inefficient groundwater extraction.

We also find that the program effect is large relative to total cost of incentives: as designed, the overall expenditure per unit of energy conserved is similar to the per-unit cost a utility company would face in procuring electricity. This suggests that a utility capable of rolling out conservation credits at low fixed cost could potentially save money if the program were carefully designed. Our program uses a combination of individual-specific benchmarks (set using verifiable baseline irrigation information) and maximum payments to avoid extreme payments for infra-marginal behavior.

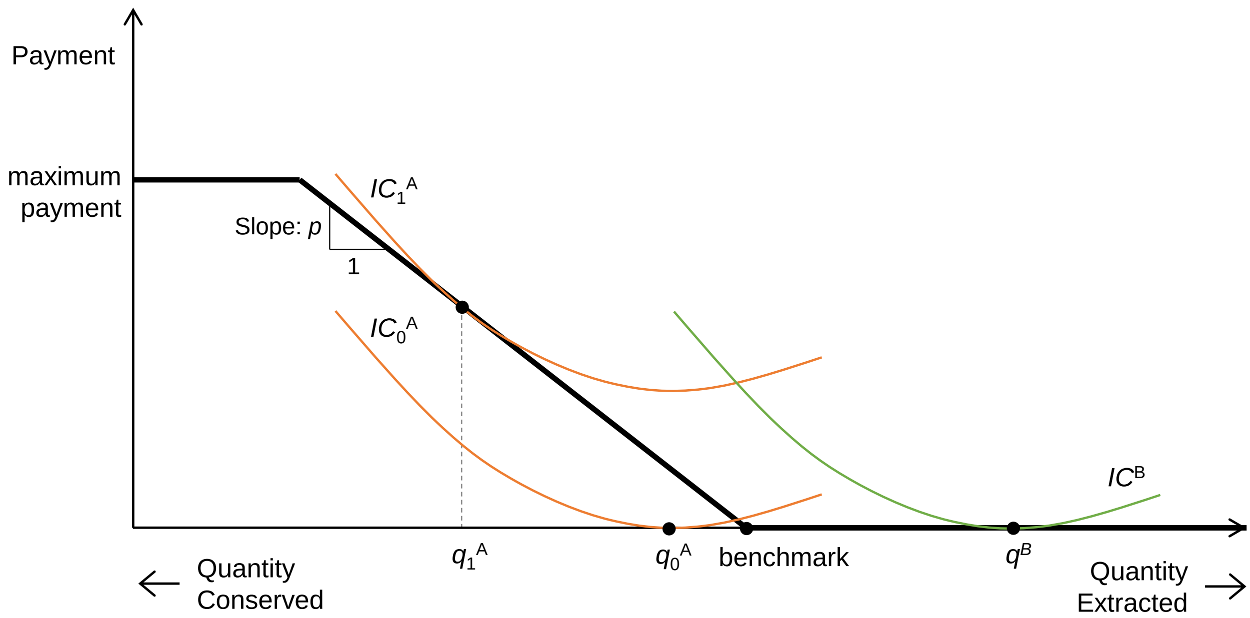


Figure 1: Budget set of conservation credits.

This figure shows the general form of the budget set created by a conservation payment contract, along with indifference curves of two example agents. The payment equals the price p times the quantity conserved below the benchmark, up to a maximum payment. Agent A is marginal and responds to the contract by reducing quantity extracted. Agent B is extra-marginal, and does not change quantity extraction in response to the program.

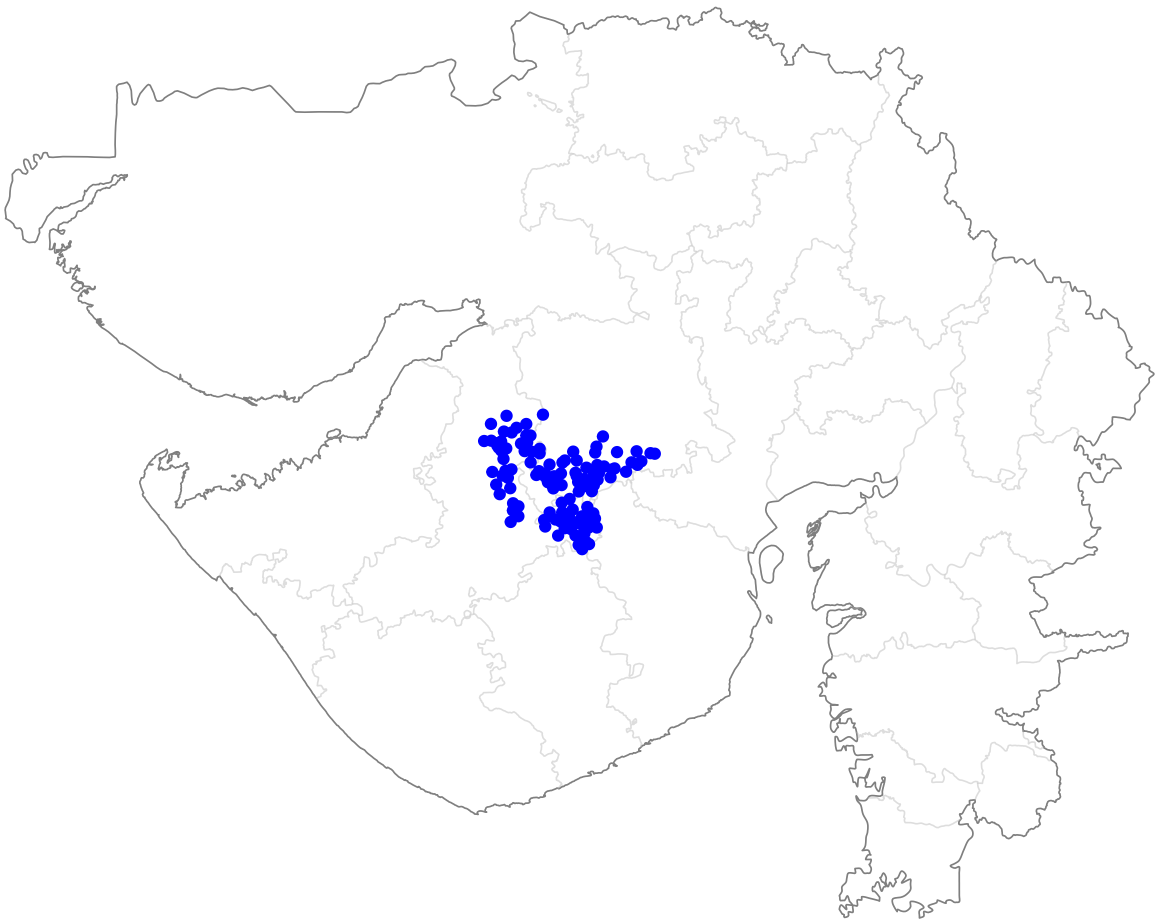


Figure 2: Villages in Study Area

Notes: This figure shows the villages in Gujarat, India where participants were enrolled as blue dots.

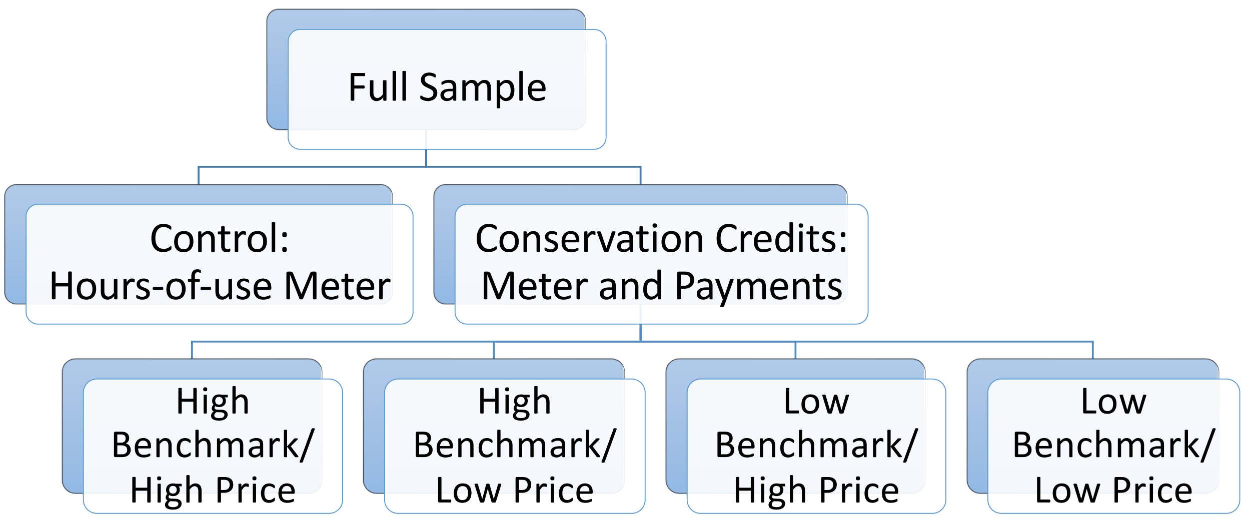


Figure 3: Intervention Design

Notes: This figure illustrates the four interventions used in the randomized experiment. Farmer sharing pools were assigned in equal proportion to the control and treatment ("Conservation Credits") groups. Within the treatment group, the four sub-treatment arms were assigned in equal proportion.

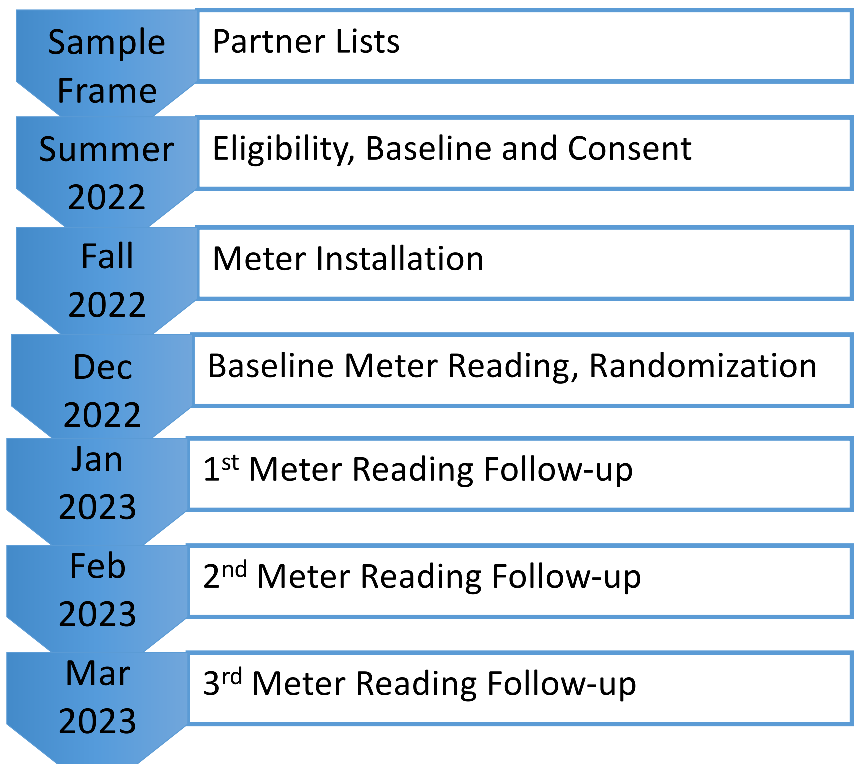
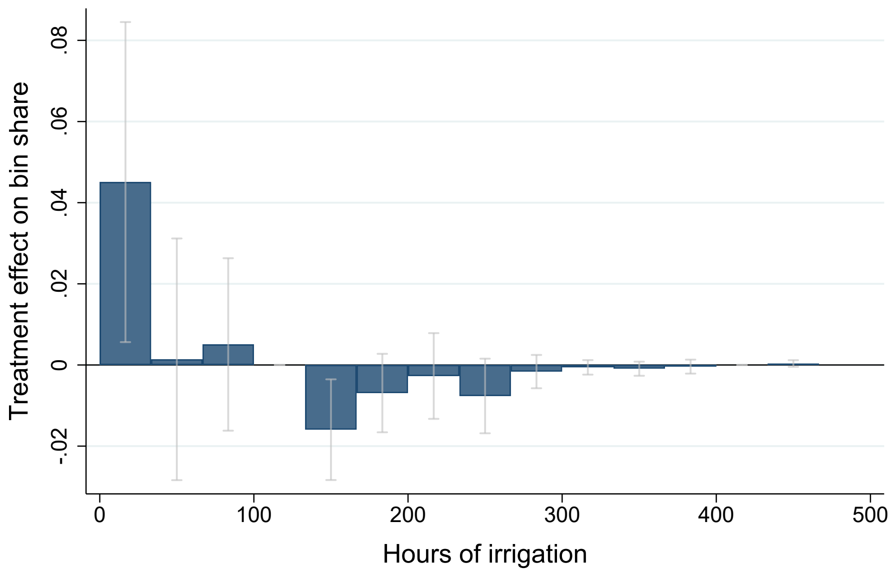


Figure 4: Experiment Timeline

Notes: This figure displays the timeline of our experimental intervention and data collection processes.

(a) Hours of irrigation



(b) Hours relative to benchmark

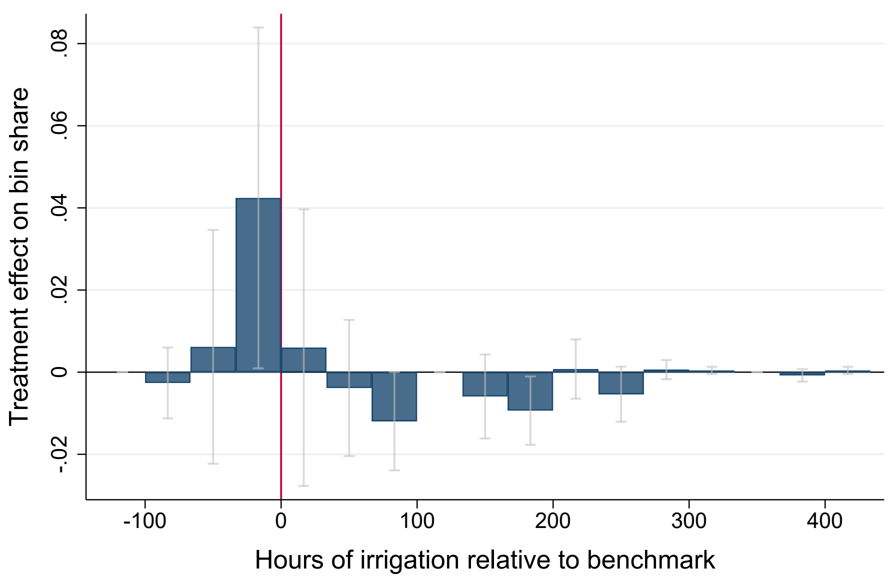
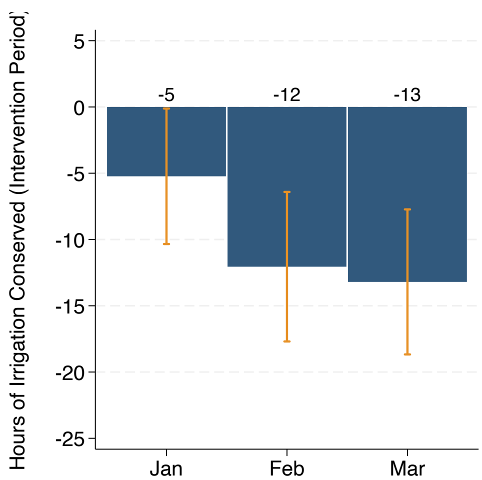
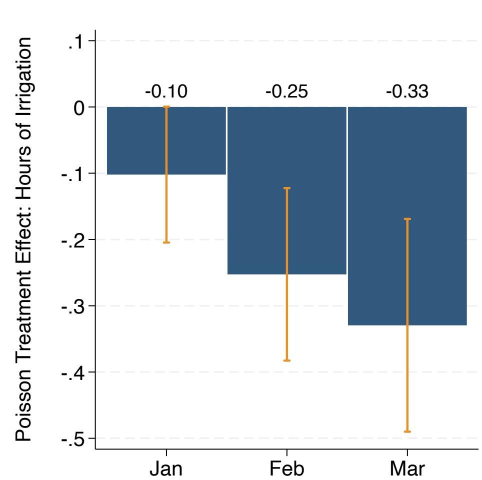


Figure 5: Bin treatment effects of payment eligibility on the distribution of irrigation hours

Notes: This figure plots estimated coefficients from regressions of binary indicators for whether a participant's value of hours of irrigation fell within specified ranges on eligibility for conservation payments. Regressions use a subset of controls. Error bars represent 95% confidence intervals.



(a) OLS



(b) Poisson

Figure 6: Treatment effects of conservation payments grew over time

Notes: This figure plots the treatment effect of eligibility for conservation payments on hours of irrigation across the three months of the intervention period. Treatment effects are estimated using double-LASSO selected controls. Error bars represent 95% confidence intervals.

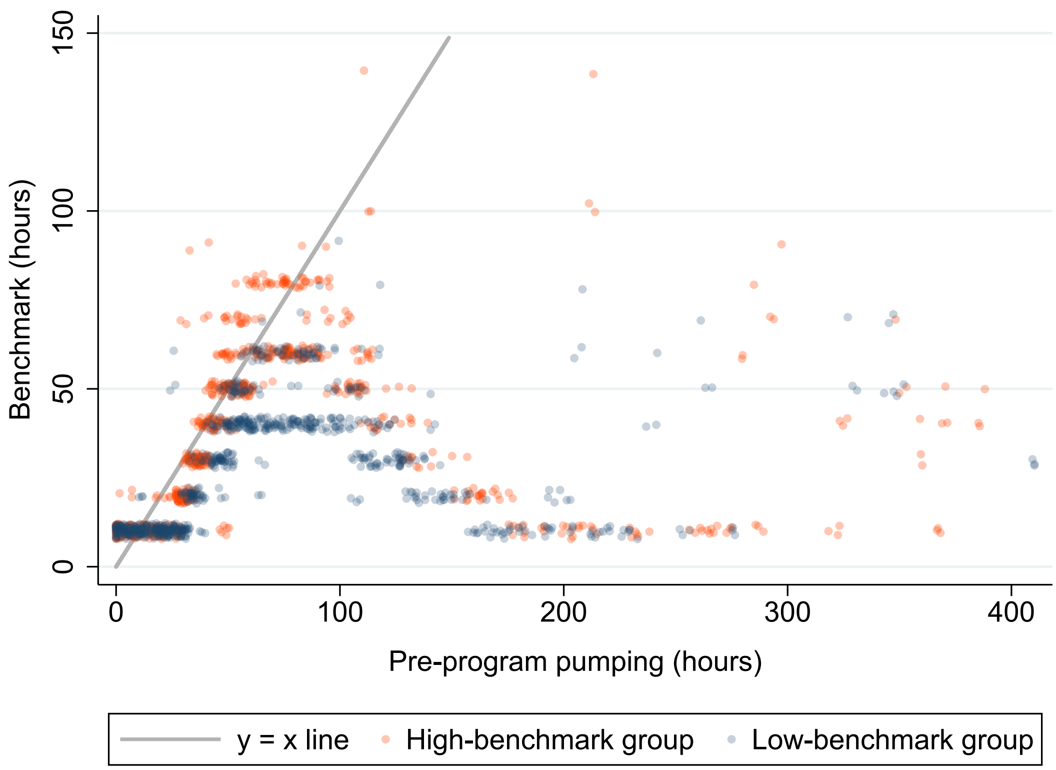


Figure 7: Benchmarks in the Treatment Group

Notes: Figure plots benchmarks set in the treatment group. Points are jittered to show density; all benchmarks were multiples of 10 hours.

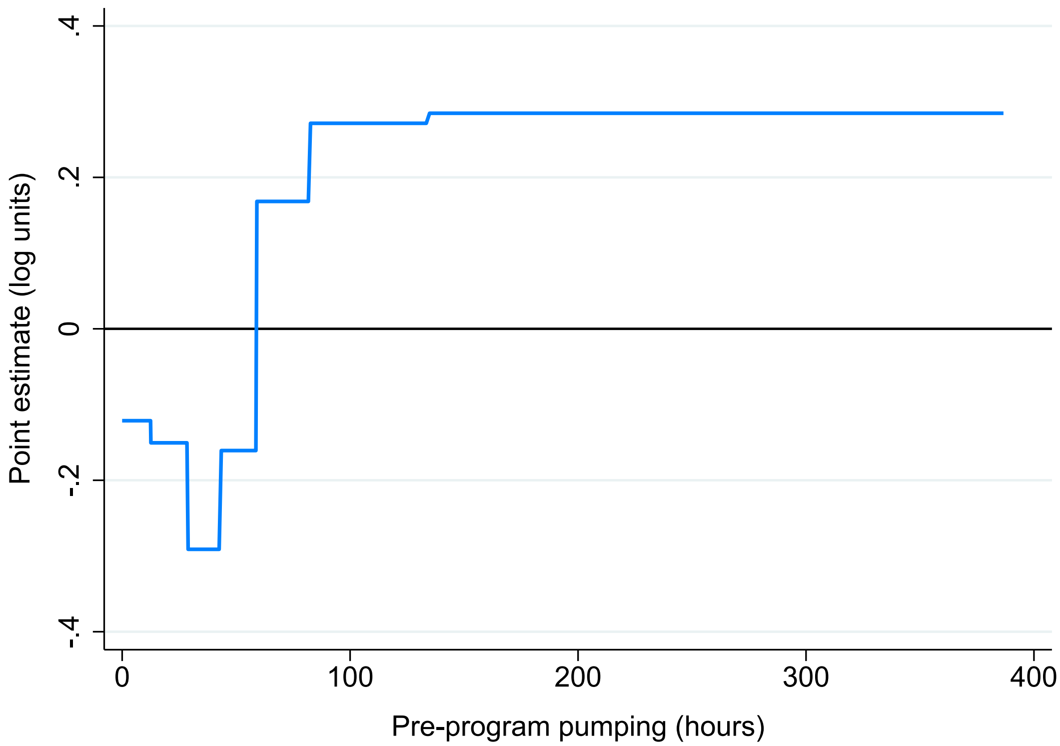


Figure 8: Effects of High Benchmarks on Pumping, By Quantiles of Pre-Program Pumping

Notes: Estimated heterogeneous effects of the randomized high-benchmark sub-treatment group relative to the low-benchmark group, by quantiles of hours of irrigation before the intervention. Figure plots coefficients in Poisson regression. Note this dimension of heterogeneity was not pre-specified.

Table 1: Baseline Summary Statistics in Full Sample and by Treatment Group

	Full Sample		Control	Treatment
	(1)	(2)	(3)	(4)
	Mean	SD	Mean	Mean
A. Demographics				
Household size	6.34	2.85	6.46	6.22
Scheduled caste/tribe or other backwards caste	0.86	0.34	0.86	0.87
Muslim	0.09	0.28	0.09	0.09
Years of education (household head)	10.94	3.39	10.88	11.00
Literacy (household head)	0.82	0.38	0.83	0.81
B. Farm statistics				
Plot hectares	1.95	1.35	1.97	1.92
Number of crops cultivated	1.96	1.08	2.01	1.91
Fraction of farmed area planted with cotton	0.53	0.41	0.54	0.53
Fraction of farmed area planted with sorghum/millet	0.15	0.25	0.15	0.16
Fraction of farmed area planted with groundnut	0.15	0.25	0.14	0.15
Fraction of farmed area planted with pulses	0.11	0.21	0.11	0.10
Has cow, ox, or buffalo	0.92	0.27	0.93	0.91
Has plow or tractor	0.50	0.50	0.50	0.50
C. Well Statistics				
Total number of active wells	1.19	0.39	1.19	1.19
Deepest well is dugwell	0.24	0.43	0.23	0.26
Deepest well is borewell	0.25	0.43	0.23	0.27
Deepest well is dug-cum-borewell	0.51	0.50	0.55	0.47
Deepest well: ever deepened	0.17	0.37	0.17	0.17
Deepest well: depth (meters)	58.62	85.17	53.66	63.12
Deepest well: max water level (meters)	16.07	36.60	14.68	17.33
Deepest well: pump power	5.61	3.27	5.46	5.75
D. Irrigation Statistics				
Pre-intervention monthly irrigation hours	71.71	71.09	69.81	73.43
Total self-reported hours of irrigation on farm	340.97	2205.91	327.45	353.25
Total self-reported hours of irrigation off farm	32.46	153.97	32.17	32.73
Purchased water for irrigation	0.01	0.11	0.01	0.01
Used drip irrigation	0.41	0.49	0.42	0.41
Used sprinkler irrigation	0.01	0.10	0.01	0.02
Used raised beds	0.69	0.46	0.69	0.68
Used rotational, strip, or zero-tillage	0.19	0.39	0.17	0.20
Used farm bunds	0.09	0.29	0.10	0.08
Test for joint orthogonality of covariates				
F-statistic			0.64	
P-value			0.93	
Sample size				
Number of individuals	989		471	518
Percent of sample	100.0		47.6	52.4

Notes: This table summarizes baseline characteristics of the sample of farmers who completed all three meter reading survey rounds during the intervention. The F-statistic and associated P-value test the joint orthogonality of all characteristics listed in the table to treatment assignment relative to the control group.

Table 2: Intent-to-Treat Impacts of Conservation Payments on Hours of Irrigation

	OLS				Poisson			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Payment Eligibility	-5.97*	-9.92***	-10.4***	-9.01***	-0.14*	-0.23***	-0.25***	-0.22***
	[3.46]	[3.21]	[2.44]	[2.92]	[0.079]	[0.073]	[0.056]	[0.067]
Payment Eligibility × High Price				-2.80				-0.047
				[3.48]				[0.084]
Control Mean	46.59	46.59	46.59	46.59	46.59	46.25	46.59	46.59
Village FE		X				X		
Lasso Controls			X	X			X	X
N Clusters	494	494	494	494	494	485	494	494
N Farmers	989	989	989	989	989	970	989	989
N Observations	2,967	2,967	2,967	2,967	2,967	2,910	2,967	2,967

Notes: The sample includes all farmers who completed all three meter reading survey rounds during the intervention. The outcome is monthly hours of irrigation by the farmer during the three intervention months (scaled to 31 days). Standard errors clustered at the randomization pair level are in brackets.

Table 3: Intent-to-Treat Impacts of Conservation Payments on Energy Use (kWh)

	OLS				Poisson			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Payment Eligibility	-23.5	-98.4	-139.8***	-83.7	-0.039	-0.17	-0.22***	-0.14*
	[70.2]	[64.5]	[40.6]	[53.9]	[0.12]	[0.11]	[0.065]	[0.082]
Payment Eligibility × High Price				-111.2*				-0.16
				[66.4]				[0.11]
Control Mean	610.81	610.81	610.81	610.81	610.81	606.16	610.81	610.81
Village FE		X						
Lasso Controls			X	X			X	X
N Clusters	494	494	494	494	494	485	494	494
N Farmers	989	989	989	989	989	970	989	989
N Observations	2,967	2,967	2,967	2,967	2,967	2,910	2,967	2,967

Notes: The sample includes all farmers who completed all three meter reading survey rounds during the intervention. The outcome is monthly kWh of energy used for irrigation by the farmer during the three intervention period months (scaled to 31 days). Energy use is calculated from hours of irrigation as described in Section 2. Standard errors clustered at the randomization pair level are in brackets.

Table 4: Cost-Effectiveness of Conservation Payments

Parameter	Value	Unit	Source
Panel A: Parameters used			
Pump motor efficiency, from a similar context	40%	-	Mitra, Balasubramanya, & Brouwer (2023)
Unit conversion constant	0.7457	kW per hp	Known constant
Mean duration of intervention	3.7	Months	Meter reading data
Panel B: Calculation of cost-effectiveness			
Average effect of program on electricity use, monthly	-139.8	kWh/month per farmer	Table 3, column (3)
Average effect of program, scaled to rabi season	-512.3	kWh per farmer	Calculated
Average conservation payments, rabi season	3369	INR per farmer	Program implementation data
Average expenditure per unit electricity conserved	6.6	INR/kWh	Calculated
Panel C: Comparisons of cost-effectiveness			
Cost of reducing electricity use through this program	6.6	INR/kWh	From above
Average cost of electricity procurement per unit sold, Gujarat	5.4	INR/kWh	Paschim Gujarat Vij Company Ltd. (2021)
Cost of electricity procurement, Punjab	7.9	INR/kWh	Mitra, Balasubramanya, & Brouwer (2023)

Table 5: Effects of Benchmarks on Hours of Irrigation, by Pre-Program Pumping

	First Stage	Reduced Form			IV	
	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	Low	High	Low	High
Payment Eligibility		-10.1***	-2.64	-20.6***		
		[2.79]	[2.61]	[6.09]		
Eligibility x High Benchmark	9.09***	4.58	-2.80	15.1**		
	[0.78]	[2.97]	[2.42]	[6.69]		
Benchmark (hours)					-0.64	1.10**
					[0.43]	[0.47]
Control Mean	0.00	46.59	18.69	85.39	0.00	0.00
Pre-program pumping (fine bin FE)	X	X	X	X	X	X
N Clusters	265	492	284	208	154	111
N Farmers	518	989	573	416	299	219
N Observations	1,554	2,967	1,719	1,248	897	657

Notes: "Low" and "High" samples split the full sample at 60 hours of first-month (i.e., pre-program) pumping. The full sample includes all farmers who completed all three meter reading survey rounds during the intervention. The outcome is monthly hours of irrigation by the farmer during the three intervention months (scaled to 31 days). Standard errors clustered at the randomization pair level are in brackets. Note this dimension of heterogeneity was not pre-specified.

Table 6: Program Cost-Effectiveness by Benchmark Group

Metric	Full sample	High-benchmark group	Low-benchmark group	Locally conservation-maximizing benchmarks
Average effect of program on pumping, rabi season (hours)	-28.7	-36.6	-24.0	-40.3
Average conservation payments, rabi season (INR)	3,369	2,605	4,157	3,482
Average expenditure per unit of pumping conserved (INR/hour)	-118	-71	-173	-86

References

Ahrens, A., Hansen, C., and Schaffer, M. E. (2018). pdslasso and ivlasso: Progams for post-selection and post-regularization ols or iv estimation and inference. *Statistical Software Components, Boston College Department of Economics*.

Asher, S. and Novosad, P. (2020). Rural roads and local economic development. *American Economic Review*, 110:797–823.

Aspelund, K. M. and Russo, A. (2025). Additionality and asymmetric information in environmental markets: Evidence from conservation auctions. Technical report, Working Paper.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. volume 31, pages 3–32. American Economic Association.

Bai, Y., Hsieh, M. H., Liu, J., and Tabord-Meehan, M. (2023). Revisiting the analysis of matched-pair and stratified experiments in the presence of attrition. *Journal of Applied Econometrics*.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain. *Econometrica*, 80(6):2369–2429.

Belloni, A., Chernozhukov, V., and Hansen, C. (2013). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650.

Blakeslee, D., Fishman, R., and Srinivasan, V. (2020). Way down in the hole: Adaptation to long-term water loss in rural india. *American Economic Review*, 110(1):200–224.

Börner, J., Baylis, K., Corbera, E., Ezzine-de Blas, D., Honey-Rosés, J., Persson, U. M., and Wunder, S. (2017). The Effectiveness of Payments for Environmental Services. *World Development*, 96:359–374.

Bruno, E. M. and Jessoe, K. (2021). Missing markets: Evidence on agricultural groundwater demand from volumetric pricing. *Journal of Public Economics*, 196.

Bruno, E. M., Jessoe, K. K., and Hanemann, W. M. (2024). The dynamic impacts of pricing groundwater. *Journal of the Association of Environmental and Resource Economists*, 11(5):1201–1227.

Burgess, R., Greenstone, M., Ryan, N., and Sudarshan, A. (2020). The consequences of treating electricity as a right. *Journal of Economic Perspectives*, 34(1):145–69.

Chakravorty, U., Dar, M. H., and Emerick, K. (2023). Inefficient water pricing and incentives for conservation. *American Economic Journal: Applied Economics*, 15(1):319–50.

Chen, J. and Roth, J. (2024). Logs with zeros? some problems and solutions. *Working Paper*.

de Chaisemartin, C. and Ramirez-Cuellar, J. (2024). At what level should one cluster standard errors in paired and small-strata experiments? *American Economic Journal: Applied Economics*, 16:193–212.

Fishman, R., Lall, U., Modi, V., and Parekh, N. (2016). Can Electricity Pricing Save India's Groundwater? Field Evidence from a Novel Policy Mechanism in Gujarat. *Journal of the Association of Environmental and Resource Economists*, 3(4):819–855.

Gonzalez-Alvarez, Y., Keeler, A. G., and Mullen, J. D. (2006). Farm-level irrigation and the marginal cost of water use: Evidence from Georgia. *Journal of Environmental Management*, 80(4):311–317.

Hendricks, N. P. and Peterson, J. M. (2012). Fixed Effects Estimation of the Intensive and Extensive Margins of Irrigation Water Demand.

Ito, K. (2014). Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing. *American Economic Review*, 104(2):537–563.

Jack, B. K. and Cardona Santos, E. (2017). The leakage and livelihood impacts of PES contracts: A targeting experiment in Malawi. *Land Use Policy*, 63:645–658.

Jack, B. K., Jayachandran, S., Kala, N., and Pande, R. (2023). Money (not) to burn: Payments for ecosystem services to reduce crop residue burning. *Working Paper*.

Jack, B. K. and Smith, G. (2016). Charging Ahead: Prepaid Electricity Metering in South Africa. *NBER Working Paper Series*.

Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Shamsudduha, M., Taylor, R. G., Fallatah, O., and Kirchner, J. W. (2024). Rapid groundwater decline and some cases of recovery in aquifers globally. *Nature*, 625:715–721.

Jayachandran, S., De Laat, J., Lambin, E. F., Stanton, C. Y., Audy, R., and Thomas, N. E. (2017). Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. *Science*, 357(6348):267–273.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102.

Lee, K., Miguel, E., and Wolfram, C. (2018). Experimental Evidence on the Economics of Rural Electrification. *Working Paper*.

Libecap, G. D. (2014). Addressing global environmental externalities: Transaction costs considerations. *Journal of Economic Literature*, 52:424–479.

McRae, S. (2015). Infrastructure Quality and the Subsidy Trap. *American Economic Review*, 105(1):35–66.

Meenakshi, J., Banerji, A., Mukherji, A., and Gupta, A. (2013). Does marginal cost pricing of electricity affect groundwater pumping behaviour of farmers? Evidence from India. *3ie Impact Evaluation Report*, 4.

Mitra, A., Balasubramanya, S., and Brouwer, R. (2023). Can cash incentives modify groundwater pumping behaviors? evidence from an experiment in punjab. *American Journal of Agricultural Economics*, 105:861–887.

Mullahy, J. and Norton, E. C. (2023). Why transform y? the pitfalls of transformed regressions with a mass at zero*. *Oxford Bulletin of Economics and Statistics*.

Pagiola, S., Honey-Rosés, J., and Freire-González, J. (2016). Evaluation of the permanence of land use change induced by payments for environmental services in quindío, colombia. *PLoS ONE*, 11.

Pai, M. M. and Strack, P. (2022). Taxing externalities without hurting the poor. Technical report, SSRN.

Pattanayak, S. K., Wunder, S., and Ferraro, P. J. (2010). Show me the money: Do payments supply environmental services in developing countries? *Review of Environmental Economics and Policy*, 4(2):254–274.

Pfeiffer, L. and Lin, C.-Y. C. (2014). The effects of energy prices on agricultural groundwater extraction from the high plains aquifer. *American Journal of Agricultural Economics*, 96(5):1349–1362.

Registrar General and Census Commissioner of India (2001). Census of India 2001.

Ryan, N. and Sudarshan, A. (2022). Rationing the commons. *Journal of Political Economy*, 130:210–257.

Saez, E., Slemrod, J., and Giertz, S. H. (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature*, 50(1):3–50.

Sekhri, S. (2014). Wells, water, and welfare: The impact of access to groundwater on rural poverty and conflict. *American Economic Journal: Applied Economics*, 6(3):76–102.

Silva, J. M. C. S. and Tenreyro, S. (2006). The log of gravity.

Smith, S. M., Andersson, K., Cody, K. C., Cox, M., and Ficklin, D. (2017). Responding to a groundwater crisis: The effects of self-imposed economic incentives. *Journal of the Association of Environmental and Resource Economists*, 4(4):985–1023.

Sovacool, B. K. (2017). Reviewing, Reforming, and Rethinking Global Energy Subsidies: Towards a Political Economy Research Agenda. *Ecological Economics*, 135:150–163.

Wolak, F. A. (2006). Residential customer response to real-time pricing: The anaheim critical-peak pricing experiment. *Center for the Study of Energy Markets Working Paper*, 151.

A A theoretical model of conservation payments

An agent receives benefit $u(q, c)$ from consumption of a resource q and a numeraire good c, where u is concave ($u_q > 0, u_c > 0, u_{qq} < 0, u_{cc} < 0, u_{qc} < 0$). The agent has income m and pays a small constant marginal cost r for each unit of the resource consumed.[21]

A principal seeks to reduce the agent's resource consumption q. They cannot change fees for use of the resource but can offer conservation payment contracts. Contracts specify a price and benchmark $\{p, b\}$ and make positive payments y that pay p for every unit of resource forgone below the benchmark (ignoring maximum payments for simplicity):

$$y = \begin{cases} p \cdot (b - q) & q < b \\ 0 & q \geq b \end{cases}$$

The agent chooses quantity q to maximize utility subject to their budget constraint:

$$\max_q u(q, c) \text{ s.t. } c + rq \leq m + y.$$

The budget constraint is piecewise, so we must consider four candidate solutions: the two corner solutions $q^2 = 0$ and $q^3 = b$, plus possible solutions on the two linear facets of the budget constraint:

$$q^0 = \arg\max_q u(q, c) \text{ s.t. } c + rq \leq m \text{ if } q^0 > b$$

$$q^1 = \arg\max_q u(q, c) \text{ s.t. } c + rq \leq m + (b - q)p \text{ if } q^1 < b$$

First-order conditions imply that q^0 and q^1 are chosen so as to set the marginal rate of substitution equal to the effective price:

$$\begin{aligned} u_q(q^0, c) &= r \cdot u_c(q^0, c). \\ u_q(q^1, c) &= (p + r) \cdot u_c(q^1, c). \end{aligned}$$

[21] In the context of groundwater, this cost might represent labor costs of irrigation, low existing volumetric fees, or internalized scarcity value.

A.1 Quasilinear utility

Let's first assume the utility function is quasilinear:

$$u(q,c) = c + v(q)$$

where $v(q)$ is strictly increasing, concave, and continuously differentiable. It has first derivatives

$$\begin{aligned} u_q(q,c) &= v'(q) \\ u_c(q,c) &= 1 \end{aligned}$$

so the FOCs become

$$\begin{aligned} r &= v'(q^0) \\ p + r &= v'(q^1) \end{aligned}$$

which we can invert to get a demand function:

$$\begin{aligned} q^0 &= D(r) \equiv v'^{-1}(r) \\ q^1 &= D(p+r). \end{aligned}$$

A.1.1 How does the benchmark affect the intensive margin?

Conditional on "being marginal," i.e., choosing the solution q^1, the benchmark does not affect quantity q^1, since it is a pure transfer. It does not appear in the demand function above.

A.1.2 How does the benchmark affect the extensive margin?

There are four possible solutions. Set aside the question of whether q^0 and q^1 exist; assume they do. When is $q^1 \succ q^0$?

$$\begin{aligned}
u(q^1;b,p) &> u(q^0;b,p)\\
c(q^1)+v(q^1) &> c(q^0)+v(q^0)\\
m+(b-q^1)p-rq^1+v(q^1) &> m-rq^0+v(q^0)\\
b &> q^1+\frac{1}{p}\left((q^1-q^0)r-[v(q^1)-v(q^0)]\right)\\
p &> \frac{1}{b-q^1}\left((q^1-q^0)r-[v(q^1)-v(q^0)]\right)
\end{aligned}$$

As shown above, the benchmark does not affect the solutions q^0 and q^1, so we can say that a higher benchmark expands the range of parameters for which this expression holds. Therefore, raising the benchmark makes the agent more likely to become marginal (choose q^1 over q^0). In other words, lowering the benchmark makes it less likely that the agent will respond to the program at all. Also, a higher price makes the agent more likely to become marginal.

When is $q^2 \succ q^1$?

$$\begin{aligned}
u(q^2;b,p) &> u(q^1;b,p)\\
c(q^2)+v(q^2) &> c(q^1)+v(q^1)\\
m+pb+v(0) &> m+(b-q^1)p-rq^1+v(q^1)\\
(p+r)q^1 &> v(q^1)-v(0)
\end{aligned}$$

The benchmark does not influence this choice. Whether or not the agent reaches the corner solution of zero depends only on demand and price, not the benchmark.

When is $q^2 \succ q^0$?

$$\begin{aligned}
u(q^2;b,p) &> u(q^0;b,p)\\
c(q^2)+v(q^2) &> c(q^0)+v(q^0)\\
m+pb+v(0) &> m-rq^0+v(q^0)\\
b &> \frac{1}{p}\left(v(q^0)-v(0)-rq^0\right)
\end{aligned}$$

Again, raising the benchmark makes the the agent more likely to choose the corner solution (choose q^2 over q^0).

A.2 CES utility

Now let's instead assume the utility function is CES, where $\alpha \in (0,1)$ and $\rho \leq 1$:

$$u(q,c) = \big(\alpha q^\rho + (1-\alpha)c^\rho\big)^{1/\rho}$$

which has first derivatives

$$\begin{aligned}
u_q(q,c) &= \big(\alpha q^\rho + (1-\alpha)c^\rho\big)^{(1-\rho)/\rho}\alpha q^{\rho-1} \\
u_c(q,c) &= \big(\alpha q^\rho + (1-\alpha)c^\rho\big)^{(1-\rho)/\rho}(1-\alpha)c^{\rho-1}.
\end{aligned}$$

The FOC for q^0 is

$$\begin{aligned}
u_q(q^0,c) &= r \cdot u_c(q^0,c). \\
\alpha q^{\rho-1} &= r \cdot (1-\alpha)c^{\rho-1} \\
\big(\frac{q}{c}\big)^{\rho-1} &= \frac{1-\alpha}{\alpha} r \\
q &= \big(\frac{1-\alpha}{\alpha} r\big)^{1/(\rho-1)} c.
\end{aligned}$$

Combining this with the budget constraint gives a demand function:

$$\begin{aligned}
q &= \big(\frac{1-\alpha}{\alpha} r\big)^{1/(\rho-1)}(m - rq) \\
q - \big(\frac{1-\alpha}{\alpha} r\big)^{1/(\rho-1)} rq &= \big(\frac{1-\alpha}{\alpha} r\big)^{1/(\rho-1)} m \\
\Big[1 + \big(\frac{1-\alpha}{\alpha} r\big)^{1/(\rho-1)} r\Big] q &= \big(\frac{1-\alpha}{\alpha} r\big)^{1/(\rho-1)} m \\
q &= \frac{\big(\frac{1-\alpha}{\alpha} r\big)^{1/(\rho-1)}}{\Big[1 + \big(\frac{1-\alpha}{\alpha} r\big)^{1/(\rho-1)} r\Big]} m \\
q &= \frac{\big(\frac{1-\alpha}{\alpha} r\big)^{-1/(\rho-1)}\big(\frac{1-\alpha}{\alpha} r\big)^{1/(\rho-1)}}{\big(\frac{1-\alpha}{\alpha} r\big)^{-1/(\rho-1)}\Big[1 + \big(\frac{1-\alpha}{\alpha} r\big)^{1/(\rho-1)} r\Big]} m \\
q &= \frac{m}{\big(\frac{1-\alpha}{\alpha}\big)^{1/(1-\rho)} r^{1/(1-\rho)} + r}
\end{aligned}$$

or,

$$q^0(p,b,m) = A^0 m$$

where

$$A^0 \equiv \Big[(\frac{1-\alpha}{\alpha})^{1/(1-\rho)} r^{1/(1-\rho)} + r \Big]^{-1}.$$

The FOC for q^1 is, by symmetry:

$$q = \big(\frac{1-\alpha}{\alpha}(p+r)\big)^{1/(\rho-1)} c.$$

And the demand function is:

$$\begin{aligned}
q &= \big(\frac{1-\alpha}{\alpha}(p+r)\big)^{1/(\rho-1)}(m+(b-q)p-rq) \\
q - \big(\frac{1-\alpha}{\alpha}(p+r)\big)^{1/(\rho-1)}(p+r)q &= \big(\frac{1-\alpha}{\alpha}(p+r)\big)^{1/(\rho-1)}(m+pb-rq) \\
\Big[1 + \big(\frac{1-\alpha}{\alpha}(p+r)\big)^{1/(\rho-1)}(p+r)\Big] q &= \big(\frac{1-\alpha}{\alpha}(p+r)\big)^{1/(\rho-1)}(m+pb) \\
q &= \frac{\big(\frac{1-\alpha}{\alpha}(p+r)\big)^{1/(\rho-1)}}{\Big[1 + \big(\frac{1-\alpha}{\alpha}(p+r)\big)^{1/(\rho-1)}(p+r)\Big]}(m+pb) \\
q &= \frac{\big(\frac{1-\alpha}{\alpha}(p+r)\big)^{-1/(\rho-1)}\big(\frac{1-\alpha}{\alpha}(p+r)\big)^{1/(\rho-1)}}{\big(\frac{1-\alpha}{\alpha}(p+r)\big)^{-1/(\rho-1)}\Big[1 + \big(\frac{1-\alpha}{\alpha}(p+r)\big)^{1/(\rho-1)}(p+r)\Big]}(m+pb) \\
q &= \frac{m+pb}{\big(\frac{1-\alpha}{\alpha}\big)^{1/(1-\rho)}(p+r)^{1/(1-\rho)} + p + r}
\end{aligned}$$

or,

$$q^1(p,b,m) = A^1 m + A^1 pb$$

where

$$A^1 \equiv \Big[(\frac{1-\alpha}{\alpha})^{1/(1-\rho)} (p+r)^{1/(1-\rho)} + p + r \Big]^{-1}.$$

A.2.1 How does the benchmark affect the intensive margin?

Assuming benchmark does not change the choice of facet/solution:

$$\begin{aligned}
\frac{dq^1}{db} &= \frac{p}{\big(\frac{1-\alpha}{\alpha}\big)^{1/(1-\rho)} p^{1/(1-\rho)} + p} \\
&= \Big[(\frac{1-\alpha}{\alpha})^{1/(1-\rho)} p^{\rho/(1-\rho)} + 1 \Big]^{-1} \\
&> 0.
\end{aligned}$$

That means raising the benchmark always increases the quantity. In other words, lowering the benchmark always increases conservation. This is because raising the benchmark is a pure income effect, so the agent wants to consume more of both goods (the resource and the numeraire).

Of course the other possibility is that raising the benchmark induces the agent to choose a solution on a different facet of the budget constraint. And the lower the benchmark, the less likely the "marginal" solution is to beat the "not participate" solution, so the quantity will increase.

A.2.2 How does the benchmark affect the extensive margin?

There are four possible solutions:

$$\begin{aligned} q^0 &= \begin{cases} A^0 m & q^0 > b \end{cases} \\ q^1 &= \begin{cases} A^1 m + A^1 pb & q^1 < b \end{cases} \\ q^2 &= 0 \\ q^3 &= b \end{aligned}$$

Set aside the question of whether q^0 and q^1 exist; assume they do. When is $q^1 \succ q^0$?

$$\begin{aligned} u(q^1; b, p) &> u(q^0; b, p) \\ \left(\alpha (q^1)^\rho + (1-\alpha) c(q^1)^\rho\right)^{1/\rho} &> \left(\alpha (q^0)^\rho + (1-\alpha) c(q^0)^\rho\right)^{1/\rho} \end{aligned}$$

If $\rho > 0$:

$$\begin{aligned} \alpha (q^1)^\rho + (1-\alpha) c(q^1)^\rho &> \alpha (q^0)^\rho + (1-\alpha) c(q^0)^\rho \\ \alpha (q^1)^\rho + (1-\alpha)(m + (b - q^1)p - r q^1)^\rho &> \alpha (q^0)^\rho + (1-\alpha)(m - r q^0)^\rho \\ \alpha (A^1 m + A^1 pb)^\rho + (1-\alpha)(m + pb - (p+r)(A^1 m + A^1 pb))^\rho &> \alpha (A^0 m)^\rho + (1-\alpha)(m - r A^0 m)^\rho \\ \alpha (A^1 m + A^1 pb)^\rho + (1-\alpha)(m + pb - (p+r)A^1 m - (p+r)A^1 pb)^\rho &> \alpha (A^0 m)^\rho + (1-\alpha)(m - r A^0 m)^\rho \end{aligned}$$

If $\rho < 0$ the inequality is reversed.

Taking the derivative of the left-hand side with respect to b:

$$
\begin{aligned}
\frac{d}{db}(LHS) &= \frac{d}{db}\alpha(A^1m + A^1pb)^\rho + \frac{d}{db}(1-\alpha)(m + pb - (p+r)A^1m - (p+r)A^1pb)^\rho \\
&= \alpha\rho(A^1m + A^1pb)^{\rho-1}A^1p + (1-\alpha)\rho(m + pb - (p+r)A^1m - (p+r)A^1pb)^{\rho-1}(1-(p+r)A^1)p \\
&= \rho p\Big[\alpha(A^1m + A^1pb)^{\rho-1}A^1 + (1-\alpha)(m(1-(p+r)A^1) + pb(1-(p+r)A^1))^{\rho-1}(1-(p+r)A^1) \\
&= \rho p\Big[\alpha(m+pb)^{\rho-1}(A^1)^\rho + (1-\alpha)\big(m + pb(1-(p+r)A^1))^{\rho-1}\big)(1-(p+r)A^1)\Big]
\end{aligned}
$$

We know that $1-(p+r)A^1 > 0$ by the definition of A^1, so this derivative is positive if $\rho > 0$ and negative if $\rho < 0$. Therefore, the left-hand side of the above inequality is increasing in b when q and c are substitutes ($\rho > 0$) and decreasing when they are complements ($\rho < 0$). So when $\rho > 0$, the inequality holds for more parameter values, and raising the benchmark makes the agent more likely to become marginal (choose q^1 over q^0). But when $\rho < 0$, the inequality is reversed, so the left-hand side decreasing in b also makes the agent more likely to become marginal. As a result, lowering the benchmark will always induce fewer agents to participate.

B Covariates included in LASSO-Selection

The following variables, which are collected prior to randomization, are fed into each double-selection LASSO after being interacted with 2 survey visit (i.e., month) indicators:

- Meter Reading
  - Average daily hours of pumping in the first month
  - Average daily energy use in the first month (calculated)
  - The natural logs of the above two variables (imputed to zero if the argument is 0)
- Baseline Survey
  - 121 Village indicators
  - Total wells on the primary farm
  - Whether deepest well has ever been deepened
  - Water level on deepest well
  - Whether deepest well went dry the previous Kharif season

- Depth of deepest well
- Indicators for deepest well being borewell or dug-cum-borewell (dugwell is omitted well type)
- Pump power for pump on deepest well
- Number of crops cultivated
- Indicator for above-median number of crops cultivated
- Fraction of farmed area planted with each of: cotton, sorghum or millet, groundnut, and pulses
- Total self-reported hours of irrigation (a) on primary farm and (b) off primary farm, previous Kharif season
- Indicator for whether purchased water for irrigation during previous Kharif season
- Indicators for typical use of raised beds, farm bunds, and low/zero-tillage practices (common water conservation practices) and for use drip irrigation previous Kharif
- Years of education (household head)
- Above-median years of education (household head)
- Indicators for Hindu and Muslim (omitted religion is other)
- Indicator for Scheduled Caste/Scheduled Tribe/Other Backwards Caste

• Groundwater Prospects Maps

  - The distance to the nearest recharge structure, fracture, escarpment, water body, observation well, mapped dugwell, mapped handwell, mapped borewell, canal, dyke, railway, stream, and road and their natural logs
  - The number of recharge structures, fractures, canals, and streams and the length of fractures in 1, 2, 3, 4 and 5 km radii
  - The fraction of land area made up of each of 42 rock types in 1, 2, 3, 4 and 5 km radii

• Satellite Images

  - The natural log of the seasonal differences in EVI and NDVI in the 2021-22 rabi season
  - the mean EVI and NDVI in November 2022

C Robustness to Attrition

Appendix Table 7 reports the response rates by treatment status and meter reading visit. While randomization was assigned at baseline, neither the farmer nor the surveyor was aware of treatment assignment until the end of the first visit shown (i.e., the "Randomization visit"). Thus, attrition at the first visit is not due to random assignment. However, following randomization, attrition increases at each subsequent meter reading visit, and differentially more so for the control group.

We examine the robustness of our results to attrition in two ways. First, we examine evidence that different types of people left the experiment in treatment and control. Second, we show that the Lee bounds on the impact of the conservation payments program on hours of irrigation do not substantially reduce our estimates.

We begin by examining the evidence on whether attrition led to baseline imbalance across treatment. Table 1 in the text shows that the treatment and control farmers are well balanced on observable baseline characteristics (p-value for joint orthogonality test is 0.93). Second, we focus on the characteristics of those who attrit after randomization. Appendix Table 8 shows that we cannot reject that these attriters' baseline characteristics are jointly orthogonal to treatment status (we also focus specifically on differential attrition by baseline irrigation hours in this table, and find no evidence on this dimension). These checks show that there are not observable differences between the treatment and control groups either among those who attrit or among those who remain in the the sample.

We next turn to the possibility that unobservable differences in attritors are driving our treatment effects. Specifically, we follow Lee (2009) in bounding the treatment effect of conservation payments on hours of irrigation under a monotonicity assumption: that no one who attrited in control would have been *more* likely to attrit in treatment. Because our outcome data are a panel, we remove the respondents in the treatment group with the highest (lowest) average hours of irrigation across the intervention period and then re-estimate Equation 3 without controls to find the lower (upper) bound of the treatment effect. We trim the treatment group by 7.2%, which is the difference in attrition rates (6.7%) between the two groups as a proportion of the retention rate between the randomization visit and the fourth meter reading in the treatment group (92%). Appendix Table 9 shows that even the upper bound on the treatment effect is negative and less than three hours different from the ITT estimate, although it not statistically different from zero (bounding also sacrifices statistical precision because rich controls cannot be included). The upper bound is correct only if differential attrition is driven by the control farmers who irrigate the fewest hours. However, Table 8 shows that baseline irrigation hours tend to be lower among retained control farmers but not among treatment farmers, suggesting that control farmers who irrigate more than average hours are driving differential attrition (baseline irrigation is strongly positively correlated with intervention period irrigation). Thus, our ITT estimates are likely to be biased upward in the positive direction. In fact, we find that the lower bound of the conservation payments impact – which is correct if the control farmers who attrit are the highest irrigators – is -19.2 hours, more than 13 hours below the ITT estimate. Overall, these bounds demonstrate that it is very unlikely that differential attrition is driving the result that conservation payments reduce groundwater consumption.

D Demand Estimation

We now use the experimental variation introduced by our program to estimate the slope of demand for groundwater irrigation. The idea is that in a program of payments for voluntary conservation, not all farmers are actually marginal to the incentive, unlike as they would be under a universal volumetric electricity price or groundwater pumping fee. Even for farmers offered payments, the marginal price is zero for those who pump for more hours than the benchmark, as well as for those who reach the maximum payment.

As a result, the treatment effect depends on specific design parameters of our program: price, benchmarks, and maximum payments. In contrast, a demand model gives us potentially more generalizable information as to how the farmers in our sample would adjust their irrigation behavior under other types of programs.

To estimate demand, we estimate instrumental variables regressions of irrigation on price, instrumenting for price with the experimental treatment groups:

$$Y_{it} = \alpha_t + \beta p_{it} + \gamma' \mathbf{X}_{it} + \varepsilon_{it}, \tag{5}$$

where $p_{it} \in \{0, 50, 100\}$ represents the effective marginal price of an hour of irrigation faced by farmer i in month t. Effective marginal price in each month is zero for control-group farmers, for treatment-group farmers who did not receive a payment, and for treatment-group farmers who reached the maximum payment. For farmers who received a payment that was less than the maximum, their effective marginal price is the price offered to them, depending on their sub-treatment group (50 or 100 INR per hour).

To boost precision while avoiding overfitting and weak instruments concerns, we use the instrumental variables LASSO method of Belloni et al. (2012). Our set of candidate instruments consists of indicators for each of the two price sub-treatments and their interactions with baseline characteristics. We again choose covariates using double-LASSO, and cluster standard errors by randomization pair.

Intuitively, IV estimates take our ITT estimates and scale them by the fraction of the sample who was in position to respond to the price incentive. Benchmarks were set too low for many farmers to reach, and too high for other farmers, such that they would have reached the maximum payment even without behavior change. The IV estimates instead attribute the full program response to the farmers for whom benchmarks were set appropriately enough to affect their behavior. This method is in the spirit of quasi-experimental estimates of the elasticity of taxable income from non-linear budget sets (as summarized by Saez et al., 2012) and of electricity demand (Ito, 2014).

D.1 Results

Table 10 reports results. Column (1) reports the first-stage relationship for an IV specification with only one instrument: overall eligibility for the conservation payments program. The estimate says that the average effective marginal price in the treatment group was 42 INR per hour.[22] This first-stage relationship is strong, with a very large F-statistic.

Column (2) shows the IV estimate with this one instrument and no covariates, while specifications in columns (3) and (4) add instruments and covariates. Moving across the table, first-stage F-statistics remain strong, while the IV estimates gain precision. Our preferred estimate is in column (4), in which both instruments and covariates are selected by double LASSO. The coefficient of -0.12 implies that average monthly irrigation hours fall by 1 hour for every 8 INR increase in the hourly price. At the middle price of 50, and the control mean of irrigation hours, this implies a price elasticity of 0.13.

One limitation of these IV estimates is that they may overstate the true price elasticity. The exclusion restriction is that the program affected irrigation only through the effective marginal price at the end of the meter reading period. This assumption will be violated if the program affected irrigation for farmers who do not end up facing positive marginal prices in a given month – for example, if they attempted to conserve below the benchmark but failed to reach their target. This is a fundamental limitation of this method for estimating demand.

However, we can still bound the price elasticity using the IV and reduced-form estimates. The IV estimate loads the entire reduced-form effect of the program onto the

[22]This value represents a weighted average of the proportion of each group that was marginal, multiplied by the price offered. We separately calculate that 58 percent of farmer-months in the sample faced a positive marginal price.

fraction of farmers with a positive effective marginal price. If some farmers change their behavior but are not observed in this group, then the true proportion of farmers affected by the program is greater than indicated by the first stage. On the other hand, it is unlikely that all farmers in the treatment group were affected by the program, so the true proportion is less than 1. The true price elasticity is then bounded above by the IV estimate, and bounded below by the reduced-form estimate: $(0.16, 0.20)$.[23]

E Supplementary Tables and Figures

[23]Scaling the ITT effect of −11 hours by the average price offered in the treatment group (75 INR per hour) gives a reduced-form effect of −0.15 hr/INR. At the middle price of 50, and the control mean of irrigation hours, this implies a price elasticity of 0.16.

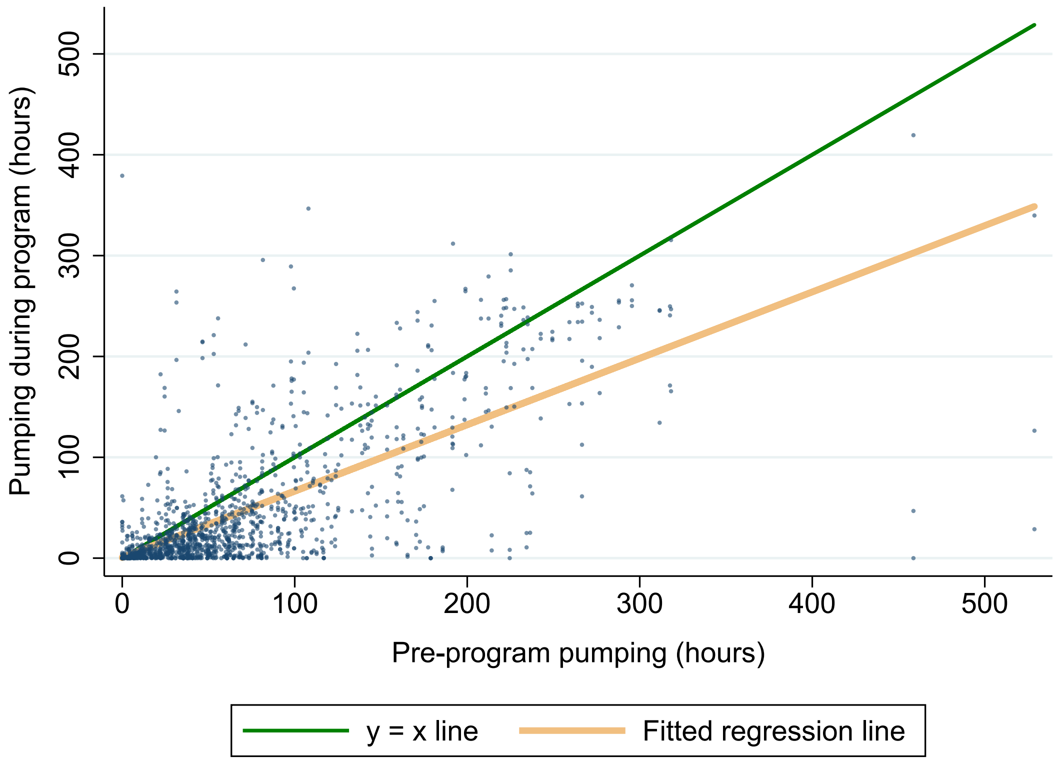


Figure 9: Predicting Pumping in the Control Group

Notes: Monthly hours of irrigation plotted against hours of irrigation before the intervention, for the control group. All values are scaled to a 31-day rate.

Table 7: Sample Retention in Treatment and Control Groups over Time

	Response rates in control group	Difference between treatment and control
	(1)	(2)
A. Response rates		
Randomization visit	0.844	0.012
	(0.364)	(0.020)
		[0.557]
Meter reading 1	0.764	0.052**
	(0.425)	(0.023)
		[0.022]
Meter reading 2	0.738	0.064***
	(0.440)	(0.023)
		[0.006]
Meter reading 3	0.722	0.067***
	(0.448)	(0.024)
		[0.005]
N	673	1308

Notes: This table summarizes attrition in the control and conservation payments treatment after baseline at each of four visits: the initial intervention visit (before randomization), and then each of three meter reading visits. The first column reports the mean and standard deviation for the control group of the farmers reached at each visit shown in the left column. Column (2) reports the estimated differences between treatment and control in the fraction of farmers reached at each visit shown in the left column. The coefficients are from a regression of an indicator from being reached on an indicator of being randomly assigned to the conservation payments treatment with no controls. Standard errors clustered at the randomization pair level are in parentheses; per comparison P-values are in square brackets.

Table 8: Baseline covariates do not predict differential retention rates in treatment and control

	(1)	(2)	(3)
Conservation Credits	0.067***	0.039	0.13
	[0.018]	[0.028]	[0.12]
Baseline irrigation hours		-0.00046*	-0.00037
		[0.00024]	[0.00027]
Baseline irrigation hours × Conservation Credits		0.00038	0.00035
		[0.00031]	[0.00036]
Control Mean	0.86	0.86	0.86
F-statistic		1.46	1.26
P-value		0.23	0.19
Baseline Controls x Treatment			X
N Clusters	541	541	540
N Farmers	1,111	1,111	1,110

Notes: This table summarizes how attrition in the control and conservation payments treatment depends on baseline covariates. Each column reports coefficient estimates from regressions of an indicator from being reached at the third meter-reading visit (i.e., being in the analysis sample) on treatment assignment. In columns 2-4, we include covariates and their interaction with treatment assignment: column 1 includes baseline survey covariate, column 2 includes irrigation hours in the month prior to randomization (available for part of the sample), and column 3 includes both baseline covariates and irrigation hours. We report the F-statistics and P-values from a test of the joint significance of all included covariates interacted with treatment assignment: in all specifications, covariates fail to differentially predict retention across treatment assignment. Standard errors clustered at the randomization pair level are in brackets.

Table 9: Lee Bounds on Impact of Conservation Payments Treatment

	ITT Estimate	Upper Bound	Lower Bound
	(1)	(2)	(3)
Payment Eligibility	-5.97*	-2.77	-19.2***
	[3.46]	[3.53]	[3.06]
Control Mean	46.59	46.59	46.59
N Clusters	495	479	471
N Farmers	989	951	951
N Observations	2,967	2,853	2,853

Notes: This table shows Lee bounds on the main treatment effect. Each column reports coefficient estimates from regressions of an indicator from being reached at the third meter-reading visit (i.e., being in the analysis sample) on treatment assignment. Standard errors clustered at the randomization pair level are in brackets.

Table 10: Demand for Groundwater Irrigation: Instrumental Variables Estimation

	First Stage	IV		
	(1)	(2)	(3)	(4)
Marginal Price		-0.14*	-0.13*	-0.12**
(INR/Hour)		[0.080]	[0.078]	[0.049]
Payment	42.2***			
Eligibility	[1.44]			
Outcome Control Mean	0.00	46.59	46.59	46.59
CD Wald F-stat		1,562.17	900.71	17.35
Fixed Effects		Month	Month	LASSO
Controls				LASSO
Instruments		Treatment	Price Sub-Treatments	LASSO
N Instruments		1	2	7
N Clusters		494	494	494
N Farmers	989	989	989	989
N Observations	2,967	2,967	2,967	2,967

Notes: The sample includes farmer-months among farmers who remained in the experiment until the final meter reading. The outcome is the monthly hours of irrigation in each of the three intervention period survey rounds. The marginal price of an hour of irrigation is instrumented using overall eligibility for conservation payments in Column 2, using all four sub-treatment arms in Column 3, and using additional high-dimensional instruments selected by double-LASSO in Column 4. Standard errors clustered at the randomization pair level are in brackets.

Table 11: Intent-to-Treat Impacts of Conservation Payments on Proxies for Crop Yields

	EVI				NDVI			
	(1) Log Diff.	(2) Diff.	(3) Mean	(4) Monthly	(5) Log Diff.	(6) Diff.	(7) Mean	(8) Monthly
Payment	0.076	0.20	-0.023	0.0089	0.066	0.0063	0.00073	0.0089
Eligibility	[0.047]	[0.17]	[0.039]	[0.026]	[0.051]	[0.0048]	[0.0034]	[0.026]
Control Mean	-0.58	0.89	1.07	0.96	-2.97	0.07	0.25	0.96
Lasso Controls	X	X	X	X	X	X	X	X
N Clusters	615	615	615	488	615	615	615	488
N Farmers	973	973	973	973	973	973	973	973
N Observations	973	973	973	2,919	973	973	973	2,919

Notes: The sample includes all farmers who completed all three meter reading survey rounds during the intervention. The outcomes include various transformations of EVI and NDVI during the 2022-23 rabi season, including: the natural log of the maximum value less the average of the first four weeks; the maximum value less the average of the first four weeks; the seasonal mean; and the monthly mean. Standard errors clustered at the randomization pair level are in brackets.

Table 12: Counterfactual Payments Under Alternative Benchmark Scenarios

	Average payment (INR per participant per month)
Actual payments, treatment group	1,238
Counterfactual payments, control group	
1. Benchmark = Pre-program value	1,872
2. Benchmark predicted from pre-program value	1,022
3. Benchmark predicted from basic characteristics	1,325
4. Benchmark predicted using rich characteristics	1,323
5. Benchmark set as in the treatment group	1,030