Over-Drilling: Local Externalities and the Social Cost of Electricity Subsidies for Groundwater Pumping*

Jesús Bueren — EUI
Xavier Giné — World Bank
Hanan G. Jacoby — World Bank
Pedro Mira — CEMFI

Abstract

Borewells for groundwater extraction have proliferated across South Asia, often encouraged by heavily subsidized electricity for pumping. Because borewells operating near one another experience mutually attenuated discharges and higher failure rates, farmers interact strategically with potentially many neighbors in deciding whether and when to drill. Using survey data from two states of southern India, we establish both the importance of this well interference externality and its influence on borewell drilling decisions. We then estimate a structural model of well-drilling as a dynamic discrete investment game played across a network of adjacent plots. Using this model, we compare the current regime of free (but rationed) electricity against an annual tax on all functioning borewells that fully defrays electricity costs. We find that the counterfactual policy, by reining in over-drilling, increases the social value of groundwater development by 39%, reducing deadweight loss by 170 US$ in present value terms per acre of land with groundwater potential, or by around 3% of its market value. We also show that taxing only newly drilled borewells at a rate 23% higher than annual electricity costs (to address the negative externalities) is nearly welfare-maximizing yet avoids a capital levy on existing well owners.

JEL codes: Q15, H23, C57
Keywords: Irreversible investment, Strategic behavior, Network games, Common property resource externalities.

*Bueren email: jesus.bueren@eui.eu, Giné email: xgine@worldbank.org, Jacoby email: hjacoby@worldbank.org, Mira email: mira@cemfi.es. We thank the World Bank Research Committee for financial support, Michael Mann and Siobhan Murray for digitizing the cadestral maps, and Gabriel Aguirre and Bernardo Ribeiro for excellent research assistance. The views expressed in this paper are those of the authors and should not be attributed to the World Bank, its Executive Directors or to the governments they represent.

1 Introduction

Millions of borewells have sprung up across the Asian subcontinent in recent decades, most of them equipped with submersible electric pumps (Shah 2010; Jacoby 2017). In India, groundwater has become the dominant source of irrigation, driving increased agricultural intensification (Jain et al., 2021) and rising rural income (Sekhri 2014). Accelerating the pace of this groundwater revolution, most Indian states currently offer farmers free or highly subsidized electricity to run their pumps, artificially inflating the economic returns to well-drilling.[1] Groundwater development has thus devolved into "drilling for subsidies", a form of rent-seeking in which smallholders sink costly wells that would not be economically viable absent these policy-induced distortions (Badiani-Magnusson and Jessoe, 2018). In this paper, we assess the social cost of electricity subsidies in the southern Indian states of Andhra Pradesh (AP) and Telangana (part of AP prior to its partition in 2014).

As a borewell operates, the water table around it drops, creating a conical draw-down region centered on the pump. If two borewells are close enough to one other, their respective cones of depression will overlap, significantly reducing the water flow from each. An additional functioning borewell nearby thus lowers the discharge of any given borewell, and may ultimately lead to its failure, i.e., when discharge is too low to warrant any cultivation whatsoever.[2] In the dry season, during which groundwater is the sole source of irrigation in most of India, well interference becomes an important, albeit highly localized, common property externality.

To be sure, even without this externality, electricity subsidies create welfare losses by encouraging over-investment, driving the gross return for the marginal borewell below the private cost of drilling plus the fiscal cost of the subsidy. Absent externalities, one could assess the social cost of over-drilling using a partial equilibrium framework, modelling the number of borewells sunk per unit area and their expected private return under a counterfactual policy of zero electricity subsidies.[3] In the presence of well interference, however, the expected private and social returns to borewell investment diverge and only a general equilibrium framework that captures both the costly externality and the resulting

[1]In 2013, Indian state governments spent US$11.4 billion to subsidize agricultural power, although this figure likely understates the fiscal drain (Sidhu et al., 2020). Since metering of usage is rare, subsidies generally take the form of low or nonexistent flat charges.

[2]The extent of the draw-down region depends on both aquifer and borewell characteristics. Well interference is accentuated in hard-rock aquifers, such as those of South India, because of the low transmissivity (velocity of horizontal groundwater flow). In our setting, pump tests conducted by the National Geophysical Research Laboratory, Hyderabad, recommend an interwell spacing of at least 250 meters to avoid interference (see Chandrakanth 2015). Blakeslee et al. (2020) detail the process of well failure in hard-rock aquifers, highlighting *local* hydrogeological features, i.e., sub-surface fractures fed from different sources of recharge. Pfeiffer and Lin (2012) discusses well interference externalities in the context of US agriculture (see also Katic and Grafton 2012 for a conceptual overview).

[3]Fafchamps and Pender (1997), for example, uses Indian panel data to estimate a dynamic model of borewell investment without externalities or strategic interactions.

strategic interactions between neighboring farmers can generate valid policy counterfactuals.

To appreciate the challenge involved, consider the related paper of Lin (2013), which estimates a strategic model of firms drilling for off-shore oil in the Gulf of Mexico. Lin obtains a tractable empirical model by restricting attention to *isolated* adjacent pairs of oil tracts leased by different companies. We cannot use this approach in our setting because any plot owner's decision to drill a borewell is potentially influenced by past and future decisions of several neighboring plot owners, who in turn are influenced by the decisions of their neighbors, and so on ad infinitum. Using (arbitrary) administrative boundaries to demarcate neighborhoods or villages within which all farmer/plots are treated symmetrically (e.g., Fu and Gregory 2019) is equally problematic in our case because the externality is highly localized.

Building on a literature surveyed in Aguirregabiria and Mira (2010), we thus develop and estimate a dynamic discrete borewell investment game played on a large (but necessarily bounded) map representing the network of adjacent agricultural plots in the locality. Structural estimation requires taking into account each plot owner's beliefs about the temporal evolution of borewells on all relevant plots, a potentially vast state-space. To avoid this "curse of dimensionality", we assume (plausibly, as we shall argue) that interference effects arise only from functioning wells located in adjacent plots *and* that decisions depend only on beliefs about such wells. The existence of a steady state in which the expected number of successful drilling attempts matches the expected number of well failures allows for a novel estimation strategy. Given model parameters, we simulate investment on the plot network map for many periods until a steady state is reached, at which point we compute beliefs based on the temporal evolution of wells in each plot owner's *adjacency*, i.e., the collection of bordering plots determining the local externality. We nest the solution to this adjacency equilibrium within a parameter estimation algorithm.

We provide two pieces of motivating evidence for our structural model. First, both borewell discharge (or flow) and failure data reveal economically significant interference effects arising from other borewells operating nearby. Second, the propensity to drill on a plot is lower, not only the more functioning borewells there are on that plot, but the more functioning borewells there are on neighboring plots. Taken together, these results establish the importance of the local externality as well as its influence on farmers' investment decisions.

Next, we pin down primitive parameters of the production technology by estimating the dynamic structural model via Simulated Method of Moments (SMM). To do so, we match the observed aggregate annual drilling rates by plot size and by the number of currently functional borewells on the plot to their model-based counterparts. As validation of our estimates, we find that model-generated data is able to quantitatively match the reduced-form impact of neighboring borewells on the propensity to drill, a moment not explicitly targeted in our structural estimation.

Based on these results, we calculate that the expected discounted present value of agricultural output minus drilling costs on land suitable for groundwater development is 77,000 Rs/acre. By contrast, the value to *society*, which differs from this private value by the expected discounted present value of electricity costs, is only 31,000 Rs/acre. We also find that charging all borewell owners fully for electricity through an annual flat tax would increase social value to 43,000 Rs/acre, or by 39%. In other words, the social cost (deadweight loss) of free electricity amounts to 12,000 Rs (170 US$) per acre, 25% of the fiscal cost of the subsidy and around 3% of farmland value in our setting.

Finally, we consider the optimal borewell tax that eliminates both rent-seeking due to free electricity and the externality costs due to well interference.[4] We compute this optimal tax under two different policies regarding extant wells. Under the first regime, all borewells (old and newly sunk) are subject to taxation but old wells can be dismantled at zero cost. In the second regime, old wells are grandfathered so that the tax only applies to new drilling (rendering it largely useless for electricity cost-recovery). We find that the socially optimal annual borewell tax is similar under both of these regimes, greater by only 1,000 Rs when taxation is restricted to newly sunk wells. Moreover, social welfare is only slightly higher under the optimal universal tax than under the similar new borewell tax. Thus, from a practical policy standpoint, grandfathering of existing borewells yields most of the social benefits of the universal tax while avoiding a (likely) politically unpalatable capital levy on existing well owners.

This paper makes both methodological and empirical contributions. Using our novel adjacency equilibrium concept, we are the first to estimate a dynamic model of irreversible investment with strategic interactions across a spatial network. Until now, the empirical literature on network games has only considered static applications (see, e.g., Acemoglu et al. 2015, Xu 2018, König et al. 2017). A static model in which all plot owners drill borewells simultaneously would not account for well failure, an important empirical phenomenon in our setting that is inherently dynamic. Moreover, a static model would be silent about the transition dynamics from the baseline scenario of over-drilling to a new steady state under a counterfactual policy regime. Analysis of transition paths is crucial to distinguishing between alternative policy treatments of existing wells (see, e.g., Domeij and Heathcote 2004 in the context of capital taxation).

We also contribute to a literature quantifying the social costs of common property externalities. Closest in spirit to our work with its focus on investment is the aforementioned oil-drilling paper by Lin (2013). Huang and Smith (2014) models boat owners' daily fishing decisions in the US shrimp industry as a dynamic game and structurally estimates the welfare cost associated with seasonal over-fishing. The key difference from our setting is that in the fishery the externality is *not* local; any shrimp caught

[4]Optimality of the centralized Pigouvian tax presupposes that landowners cannot restrain socially undesirable drilling through side-payments to neighbors. Aside from enforcement issues, one argument against such a Coaseian solution in our setting is the complex, multi-lateral, negotiations it would require across the entire network of agricultural plots.

by one boat reduces the potential catch of all other boats equally. Regarding groundwater pumping from existing wells, Sears et al. (2022) estimates a dynamic model of strategic interactions among neighboring extractors in California, but eschews the spatial network structure that is central to this paper. In the Indian context, Ryan and Sudarshan (2022) estimates the welfare cost associated with over-pumping in the northern state of Rajasthan, without considering well interference, focusing instead on the non-local, aquifer-wide, externality.[5] Ryan and Sudarshan (2022) also takes the number of borewells as given, thus ignoring drilling costs (and over-drilling), our primary concern. It finds that electricity rationing to agriculture leads farmers in Rajasthan to pump roughly the socially optimal quantity of groundwater on average. Similar rationing in our study areas of AP and Telangana undoubtedly also limits over-pumping, rendering this intensive margin distortion small in comparison to the extensive margin distortion that we emphasize.

The rest of the paper is organized as follows. In Section 2, we describe our setting and data. Section 3 presents panel data estimates of the determinants of drilling along with a joint econometric model of well flow and failure that accounts for unobserved heterogeneity. Section 4 lays out the dynamic structural model of borewell investment while Section 5 discusses the SMM estimation algorithm and presents the results. The analyses of counterfactual policies and the optimal borewell tax follow in Section 6. Section 7 concludes.

2 Setting and Data

2.1 Context

Before its partition, Andhra Pradesh was one of the most important agricultural states of India, accounting for about 7 percent of gross cropped area nationally, with groundwater supplying roughly half of its irrigation. As argued by Kumar et al. (2011), however, the economic efficiency of groundwater extraction in AP has been falling substantially, with the tripling in the number of borewells to more than 1.5 million from 1995-2010 (see Jacoby 2017), leading to high rates of well failure, lower area irrigated per well, and higher energy requirements for groundwater pumping due to well interference. Power supply to agriculture for running electrical pumps has, meanwhile, become a political issue all over India. In 2004, a newly elected government in AP abolished flat rate electricity charges to farmers, which had previously covered just 11 percent of the cost of provision, and introduced free agricultural

[5] Well interference implies that Ryan and Sudarshan (2022) likely understates the social cost of extraction. It also ignores revenue from groundwater sales to neighboring farmers. To avoid the latter complication in our study, we deliberately chose a setting in which groundwater markets are virtually nonexistent (see Giné and Jacoby 2020).

power, a move swiftly followed by the major states of Tamil Nadu, Karnataka, and Punjab.[6] Farmers in AP and Telangana typically run their pumps continuously during the fixed number of hours (7-9) per day when this free electricity is made available for agricultural use (Fishman et al., 2023).

Much of South India is underlain by shallow hard rock aquifers with limited groundwater storage capacity. Recharge from monsoon rains is thus largely depleted through pumping during the subsequent dry season; there are no deep groundwater reserves available to 'mine'.[7] As seen in Figure 1, the time-series of depth to water table across AP, a measure of overall resource depletion, is dominated by the intra-annual variability, showing practically zero trend from 1998-2014, the most recent years for which we have consistent data before the partition.[8] Well interference, therefore, is the predominant groundwater pumping externality in our setting, one that is both localized and *static*, affecting only current groundwater availability.

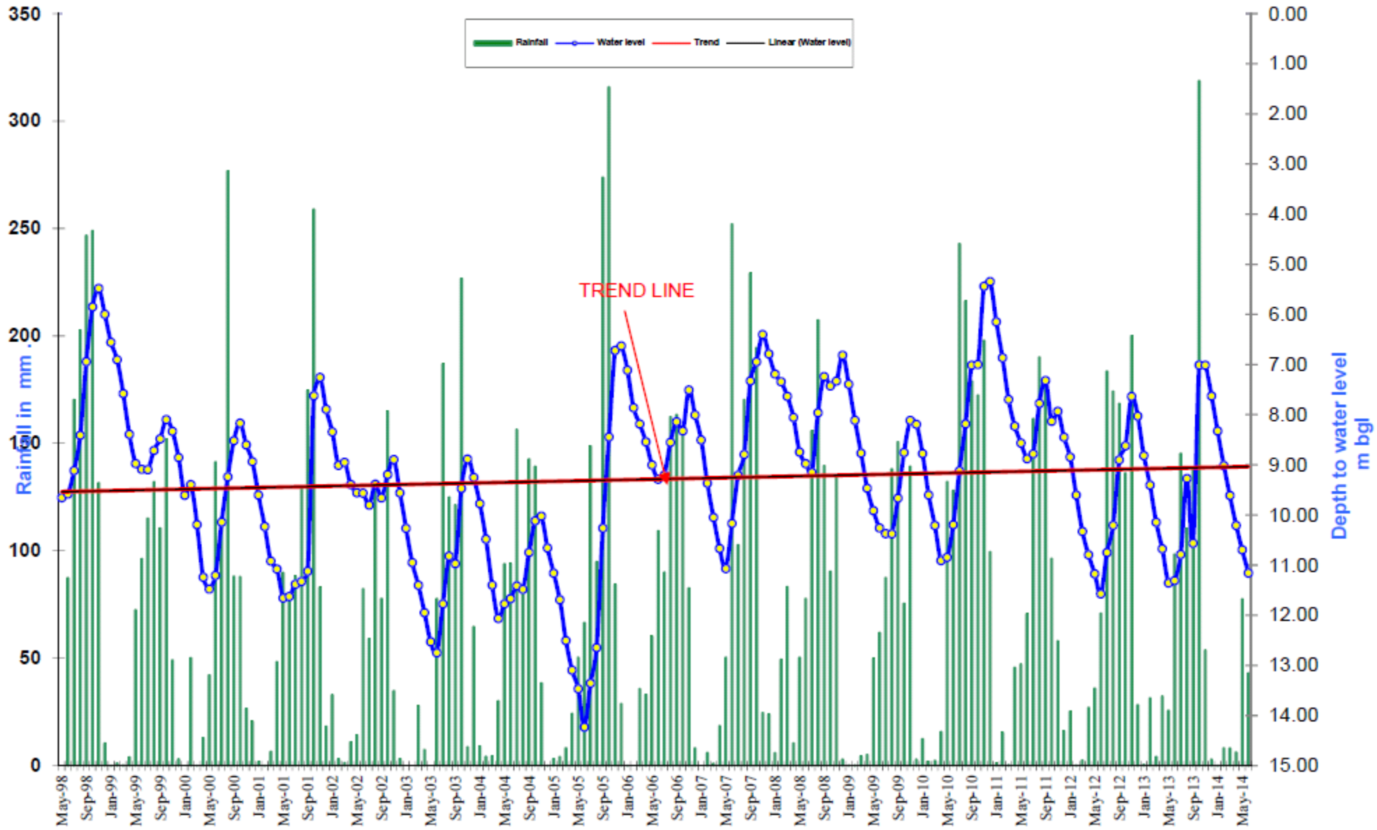
Our data come from the drought-prone districts of Anantapur (Andhra Pradesh) and Mahabubnagar (Telangana), originally the backdrop for the weather insurance study of Cole et al. (2013). As shown in Giné and Jacoby (2020), groundwater availability and the related development of groundwater markets in these drought-prone districts is limited compared to districts in the intermediate range of annual rainfall and, especially, to those in water-abundant coastal AP. Only farmers with access to a functioning borewell can cultivate during the dry (rabi) season, typically growing groundnut, maize, mulberry, and paddy in Ananatapur and paddy and groundnut in Mahabubnagar. In the wet (kharif) season, during which groundwater is used to supplement monsoonal rainfall, the main crops in both districts are paddy, sorghum, and groundnut. Lastly, a feature of our setting contributing to well interference externalities is the high degree of landownership fragmentation. To obtain the typical spatial layout of separately owned plots, we digitized cadastral maps for at least one village in each of the 12 mandals (sub-districts or counties) represented in our survey data (see Appendix D). In all, we digitized 14 maps containing 12,330 land parcels. The median plot size is only 2.02 acres.

[6]Shah et al. (2012) estimates that these subsidies in AP amounted to 94% of the gross value of its agricultural output before partition. The corresponding figure in the more agriculturally productive state of Punjab is only 12%. Note that Shah et al. (2012) uses an annual electricity cost per borewell of about US$450 for the entire state of AP circa 2010, whereas we obtain a much more conservative figure of US$180 (8,500 Rs) in our study areas (see Appendix C).

[7]By contrast, water-mining is a major concern in the deep alluvial aquifers of northwest India (see Fishman et al. 2011; Sayre and Taraz 2019; Ryan and Sudarshan 2022).

[8]Hora et al. (2019) argues that such water table trends are biased upward by relying on surviving (i.e., non-failed) observation wells to measure groundwater levels across time. Indeed, our analysis of well failure in Appendix E is consistent with a secular, but rather slow, decline in water tables in our study area.

Figure 1: Water table fluctuations: 1998-2014

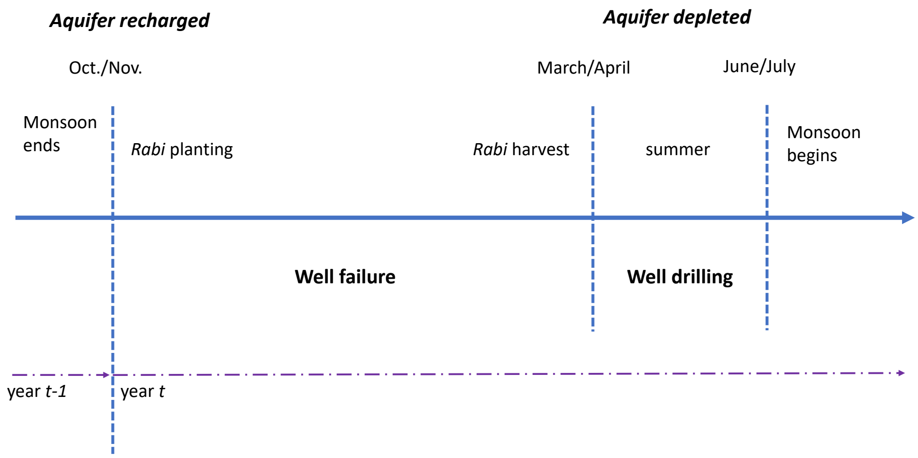


Notes: Average depth to water table in meters below ground-level from all state observation wells and rainfall in millimeters by month (Source: Government of Andhra Pradesh, Groundwater Department, 2014, http://apsgwd.gov.in/swfFiles/reports/state/monitoring.pdf).

2.2 Representative plot sample and retrospective panel

Our (second stage) structural estimation described in Section 5 is based on a representative sample of plots from Anantapur and Mahabubnagar districts. In 2017, we were able to re-interview 1,436 of 1,488 randomly selected farm (landowner) households originally surveyed by Cole et al. (2013). The 2017 survey instrument included a history of well-drilling attempts on and around each of the household's plots since 2011 as well as information on every borewell present on each plot regardless of whether currently functional (i.e., having non-negligible discharge). Using this information, we construct a retrospective five-year panel with information on drilling attempts and the number of functional borewells, as described below, for 2,862 plots. Highlighting the representativeness of these data, median plot area in this sample is 2.00 acres, virtually the same as that found in the independent plot sample from the 14 digitized cadastral mentioned above.

Figure 2: TIMELINE

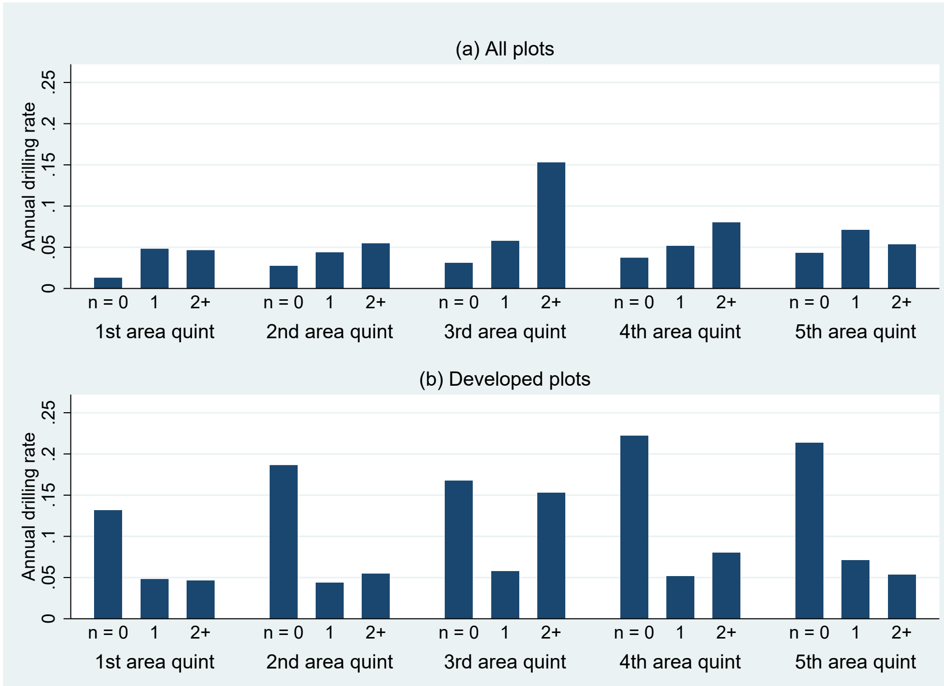


Timeline Figure 2 provides an event timeline to guide our empirical and theoretical analyses. The "year" begins with rabi season planting just after the monsoon. Borewell drilling occurs in the pre-monsoon (summer) season when water tables are at their lowest, thereby assuring farmers that, if successful, the new borewell will yield groundwater throughout the rabi season. New borewells are thus available for pumping only in the year following a successful drilling attempt, with year t "success" defined as being functional at least during year $t+1$. Since our survey does not record the exact month that a borewell failure occurs (or is "realized" by its owner), we assume that, if a borewell is reported as failed in year t, this failure is effective only as of the beginning of year $t+1$, reducing output from year $t+1$ forward; in other words, we treat this borewell as functioning throughout year t. For the plot-level retrospective panel, we drop data from 2017 because, as the survey was administered in May, there is the potential for truncation of well drilling attempts and failures occurring later in 2017.[9]

Groundwater development status Most plots in our representative sample show no drilling activity over the retrospective panel nor in prior years as evidenced by the absence of borewells. Our interpretation is that such plots are *permanently* unsuitable for groundwater development and that, as a result, the model of Section 4 does not apply to their owners. We thus introduce two plot types, "developed" and "undeveloped" and, using the retrospective panel, assign developed indicator $\mathcal{D} = 1$ to a plot if there was ever a functioning borewell or a drilling attempt made on it from 2012-16; otherwise

[9]For consistency with the adjacency survey panel (see below), which we use for a crucial model validation exercise, we also drop data on drilling activity in 2011.

Figure 3: Drilling attempts



Notes: Average annual drilling attempt rate by plot area quintiles and by number (n) of functioning borewells on the plot in the previous year. Panel (a) is for representative sample of plots; panel (b) is only for developed plots (those which had any functioning borewell or drilling attempt during 2012-16).

$\mathcal{D} = 0$.[10] By this measure, 1,108 plots (38.7% of the representative sample) are developed. In Appendix B, we show that, the propensity to develop a plot is strongly related to local hydrological features as reflected in the pre-sample subjective probability of drilling success.

Drilling patterns There were 526 drilling attempts made on 437 plots between 2012-16, only 197 (37.5%) of which were successful. Panel (a) of Figure 3 shows annual drilling rates by the number of functioning borewells and plot area quintiles. While drilling is least likely to occur in plot-years during which there are no functioning borewells ($n = 0$), this merely reflects the fact that many such plots are undeveloped. Focusing on developed plots, panel (b) reveals that plot-years with $n = 0$ actually have the *highest* drilling rates, consistent with diminishing marginal returns. This difference between panels (a) and (b), as well as the greater drilling when $n > 1$ than when $n = 1$ in panel (a), highlights the importance of heterogeneity; where drilling is profitable, there is both more drilling *and* more functioning borewells. Finally, Appendix B shows that there is no association between the pre-sample

[10]Since we do not observe the entire borewell/drilling history on a plot, $\mathcal{D}$ is a noisy measure of permanent development status, but given the five-year time window the measurement error is arguably small.

wealth of the plot owner and the propensity to drill *conditional* on the plot being developed. This and other descriptive evidence adduced in Appendix B suggests that it is reasonable to abstract from financial frictions in the borewell investment model of Section 4.

2.3 Adjacency survey and retrospective panel

Our analyses of the determinants of drilling (Subsection 3.3) and of borewell failure (Subsection 3.4) rely on an adjacency survey that was part of the 2017 instrument. An *adjacency* is defined as the set of all agricultural plots contiguous to the reference plot, inclusive of it. We assume that only other borewells operating in the adjacency create interference effects on the reference plot, imposing a negative externality. Put differently, borewells outside the adjacency do not influence the flow and failure of borewells in the reference plot. Given the typical size of plots and the range of well interference effects in our setting (Chandrakanth, 2015), we believe that this assumption is sensible.[11]

All 1,057 farmers with an eligible reference plot were administered an adjacency survey. A plot was eligible if at least one drilling attempt had been made in the last seven years either on the plot itself or within a 500 meters radius of the plot (if the household had two or more eligible plots, one was chosen at random). The adjacency survey asks each reference plot owner for retrospective information about the existence and status (functioning or not) of all borewells in the adjacency over the previous seven years. Following our timing conventions, we match drilling activity and borewell failure on reference plot i in year t with the number of functioning wells on the reference plot, n_{it}, and with the number of functioning wells in the adjacency outside of reference plot, $\mathcal{N}_{it}$, both observed at the *beginning* of year t (i.e., before any year t failures). We drop data from 2011 because we do not have $\mathcal{N}_{it}$ for that year.

Throughout the paper, we denote the total number of functioning wells in the adjacency by

$$N_{it} \equiv \mathcal{N}_{it} + n_{it}. \tag{1}$$

Arguably, respondents may recall the status of borewells on adjacent plots less accurately than those on (their own) reference plots. Thus, in our econometric analyses in Section 3, we allow $\mathcal{N}_{it}$, but not n_{it}, to be measured with error; specifically, misclassified as functioning (when actually failed) or as failed (when actually functioning).

[11] In a chessboard configuration of identical plots averaging one hectare (as in our data) with borewells located at the center, the distance between a borewell in the reference plot and one elsewhere in the adjacency would be 100-140 meters, well within the range of interference effects mentioned in footnote 2. Expanding the definition of adjacency to include a second ring of identical plots would increase the average distance between wells to 200-280 meters, which is beyond the range for interference effects in our setting.

2.4 Borewell failure panel

To estimate the annual probability of well failure, we use the adjacency survey to construct a 2012-16 panel of reference plot borewells that are *at risk* of failure. Wells enter the failure panel in the year after they are sunk; we drop those sunk in 2016 because they would not be at risk of failure until 2017. Since failure is an absorbing state, a well exits the panel in the year following its failure. For reasons discussed in Section 3.4, when the reference plot has multiple functioning wells we only include in the panel the oldest, i.e., the first well sunk. The result is an unbalanced failure panel of 697 borewells over as many as five years; 320 of the 1,057 adjacencies do not contribute to the panel as they have no functioning borewells on the reference plot over the sample period. Of the 606 borewells that were functional going into 2012, about a third (195) had failed by 2016 leading to an average annual failure rate of 7.3% (see Appendix Table A.1).[12]

A key issue in modelling well failure is duration dependence, as the probability of failure depends on the age of the well. If water tables were trending downward, then older and thus shallower wells would dry up first. With a non-constant hazard rate of well failure, farmers would profitably take into account not only the number of adjacent functioning borewells but also their ages, increasing the state space and thus introducing considerable complexity into the structural model. While Figure 1 suggests that water tables in our setting have been fairly stable over the last two decades, we assess the importance of duration dependence by focusing on extant borewells in 2012, when they had a median age of 12 years. Regressions reported in Appendix E reveal a significant positive association between well age and subsequent failure, but this effect is entirely driven by the fewer than 10% of borewells that were more than 20 years old in 2012. For investment planning purposes, then, and given discounting, it is reasonable to assume that farmers view the well failure hazard as essentially constant.

2.5 Borewell flow panel

Data on discharge (well flow) were collected in the 2010 survey from all functioning borewells for the 2009-10 rabi season and, in the 2017 survey, for the 2016-17 rabi season. Farmers were asked to assess flow at both the beginning and end of the rabi season based on the fraction of the outlet pipe that was full when pumping water (see Giné and Jacoby 2020), so the flow measure varies between ("minimal" coded as) 0.1 and ("full" coded as) 1.0, with one-quarter, one-half, and three-quarters flow in between. Reflecting the cyclicality of water tables during the rabi season seen in Figure 1, average flow assessments in 2016-17 (2009-10) fall from 0.84 (0.62) at the start of rabi to 0.57 (0.35) at the end. We focus here on end-of-season flows, since well interference becomes more salient as the local aquifer is drawn down.

[12]Blakeslee et al. (2020) also report high rates of well failure in the neighboring South Indian state of Karnataka.

To estimate flow probabilities in Section 4, we construct a balanced 2-year panel of 514 functioning wells present in both the 2010 and 2017 household/plot surveys. We find that average end-of-season flow declined in the panel by almost half from 2010 to 2017 (see Appendix Table A.2), which may be attributable to differences in the respective monsoons. According to data on total precipitation from June to November from our study area (see Appendix Figure A.1), rainfall during the 2016 monsoon season, responsible for rabi 2016-17 recharge, fell roughly 30% short of that in 2009.

2.6 Subjective drilling success

Both the 2010 and 2017 household/plot surveys ask, "if anyone tried to dig a well today within 500 meters of this plot, what do you think is the percent chance that the person would succeed?" A significant advantage of these data relative to those on actual drilling success mentioned above is that they provide a subjective probability around each plot regardless of whether there was ever an attempt to drill on it. Moreover, it is well known that, due to cognitive biases, subjective probabilities underlying real-world behavior may deviate from objective probabilities (e.g., Armantier and Treich 2009). Because subjective assessments of drilling success could be affected by recent drilling activity, however, we only use responses from the 2010 survey, predating our 2012-16 drilling data, in our analyses.

Consistent with drilling occurring when well interference effects are negligible and when groundwater levels are at their annual nadir (i.e., in the summer), our structural estimation assumes that the drilling success probability is independent of both the number of functioning wells in the adjacency and of monsoon rainfall. Finally, since, by assumption, our theoretical model of Section 4 applies only to developed plots, we use the subjective drilling success probabilities for developed plots as primitives in our structural estimation. The average success rate on developed plots is 0.520, higher than the actual success rate of 0.375 for all drilling attempts from 2012-16 (see Appendix Figure A.2 for more details).[13]

3 Preliminary Estimation

In this section, we first develop the econometric tools for our analysis, then turn to the determinants of borewell drilling to assess the empirical relevance of well interference, and, lastly, present estimates of the flow-state and well failure processes used as primitives in the structural model.

[13]The relatively low success rates, whether subjective or actual, reflect the hydrology of shallow hard-rock aquifers, where groundwater storage is highly localized, confined to certain narrow fissures. By contrast, drilling success in a deep alluvial ("bathtub"-type) aquifer would be virtually assured.

3.1 Unobserved heterogeneity

To allow for time invariant unobserved heterogeneity in the drilling and flow/failure empirical models, we specify the probability of an outcome as a function of an index z_{it} for plot i in period (year) t as

$$z_{it} \quad = \quad \nu_i + \beta_0 + \beta_1 N_{it} + \beta_2 R_{mt-1} + \varepsilon_{it}, \tag{2}$$

where ν_i is a normally distributed random effect, N_{it} is the number of functioning wells in the adjacency at the beginning of the period, and R_{mt-1} is a dummy variable that takes a value of one if monsoon rainfall in mandal m in year $t-1$ (the past monsoon) exceeds the 2009-17 average for the 12-mandal study area as a whole. The time-varying error ε_{it} is assumed iid logistic.

A key concern is that ν_i may be correlated with (N_{it}, R_{mt-1}). However, since nonlinear probability models do not lend themselves to fixed effects approaches (except in some special cases), we employ *correlated random effects* (CRE). As discussed in Wooldridge (2016), the validity of CRE depends on N_{it} and R_{mt-1} being strictly exogenous conditional on ν_i.[14] In particular, let

$$\nu_i \quad = \quad \gamma_1 \bar{N}_i + \gamma_2 \bar{R}_m + \mu_i, \tag{3}$$

where bars denote reference plot-specific means of the corresponding explanatory variable and μ_i is a continuously distributed mean zero random effect. Substituting into (2) yields

$$z_{it} = \mu_i + \beta_0 + \beta_1 N_{it} + \beta_2 R_{mt-1} + \gamma_1 \bar{N}_i + \gamma_2 \bar{R}_m + \varepsilon_{it}, \tag{4}$$

which is the index function that we use in our estimations below.

3.2 Misclassification error

The reported number of functioning wells on *neighboring* plots, $\mathcal{N}_{it}$, may be subject to recall error, leading to a non-classical form of measurement error.[15] We thus assume, quite plausibly, that the number of *existing* neighboring wells $\mathcal{N}_{it}^E$ is accurately observed. Ignoring n_{it}, which is accurately measured by assumption, we thus want to estimate the probability that some outcome Y_{it} depends upon the true number of functioning wells in the adjacency $Pr(Y_{it}|\mathcal{N}_{it})$. Although $\mathcal{N}_{it}$ is not perfectly

[14]That is, given ν_i, N_{it} and R_{mt-1} must be uncorrelated with all present and future values of ε_{it}. In the case of drilling, however, strict exogeneity is violated if we condition on $N_{it} = \mathcal{N}_{it} + n_{it}$ (or on $\mathcal{N}_{it}$ and n_{it} separately) because a successful drilling attempt in period t will augment n_{it} by one. Section 3.3 discusses how we address this issue.

[15]While there has been recent progress in the econometrics literature on models of misclassification (e.g., Mahajan 2006; Hu 2008), no tractable general approaches exist applicable to our specific situation.

observed, we know that

$$Pr(Y_{it}|\mathcal{N}_{it}^E) = \sum_{k=1}^{\mathcal{N}_{it}^E} Pr(Y_{it}|k)Pr(k|\mathcal{N}_{it}^E), \tag{5}$$

where $Pr(k|\mathcal{N}_{it}^E)$ is the discrete probability density of the true number of functioning wells outside of the reference plot. This density is of the binomial form

$$Pr(k|\mathcal{N}_{it}^E) = \frac{\mathcal{N}_{it}^E!}{k!(\mathcal{N}_{it}^E - k)!} p^{\mathcal{N}_{it}^E - k}(1-p)^k, \tag{6}$$

where p is the underlying annual probability of well failure.[16] The misclassification error model (MEM) estimator then assumes that the likelihood contribution conditional on unobservable μ is

$$\ell_i^y(\mu) = \prod_{t=1}^{T} \sum_{k=1}^{\mathcal{N}_{it}^E} Pr(Y_{it}|z_{it}(k,\mu))Pr(k|\mathcal{N}_{it}^E, \hat{p}). \tag{7}$$

3.3 Determinants of drilling

Our drilling logit is based on CRE-MEM likelihood (7) assuming that μ is normally distributed. A successful drilling attempt in period $t-1$, however, increases the number of functioning wells on the reference plot by one, rendering n_{it} endogenous conditional on μ_i. Wooldridge (2005) suggests controlling for the initial conditions in such cases, altering equation (4) as follows:

$$\begin{aligned} z_{it}^d &= \mu_i + \beta_0 + \beta_1 \mathcal{N}_{it} + \beta_2 \mathbb{1}_{n_{it}=1} + \beta_3 \mathbb{1}_{n_{it}>1} + \beta_4 R_{mt-1} \\ &+ \gamma_1 \bar{\mathcal{N}}_i + \gamma_2 \mathbb{1}_{n_{i1}=1} + \gamma_3 \mathbb{1}_{n_{i1}>1} + \gamma_4 \bar{R}_m + \gamma_5' Z_i + \varepsilon_{it}. \end{aligned} \tag{8}$$

In addition to allowing for separate effects of own (β_2 and β_3) and neighboring (β_1) borewells, equation (8) includes the vector of controls Z_i consisting of log reference plot area, number of plots in the adjacency, and mandal dummies. Well interference induces strategic substitutability between drilling decisions of neighboring plot owners, implying that $\beta_1 < 0$. Since $\mathcal{N}_{it}$ reflects only *past* investment behavior by neighbors, β_1 in equation (8) is identified even if contemporaneous plot-specific drilling shocks ε_{it} are (spatially) correlated between plots in the same adjacency.[17]

Table 1 reports the determinants of annual drilling decisions on 763 developed plots out of the

[16] For simplicity, we take p to be constant over time and across adjacencies in each district. From our well failure data, we estimate $\hat{p} = 0.104$ in Anantapur and $\hat{p} = 0.052$ in Mahabubnagar.

[17] While identification does break down if ε_{it} is both spatially *and* serially correlated (a version of Manski 1993's reflection problem), a spurious finding of strategic substitutability could only be explained by either negative spatial or negative serial correlation in drilling shocks, either of which is implausible. In a more likely scenario of positive spatial and serial correlation of drilling shocks, our estimate of β_1 would be biased upward and thus *away* from strategic substitutability.

1,057 reference plots covered by the adjacency survey. The average annual drilling rate on this sample is 0.097. In column 1, a CRE logit specification that ignores misclassification error shows a negative but insignificant effect of $\mathcal{N}_{it}$ on drilling (first row). After correcting for misclassification using the MEM estimator, we find (column 2) that having more functioning wells in the adjacency outside of the reference plot *does* significantly reduce the likelihood of further drilling.[18] This is strong evidence of strategic substitutability between neighbors' drilling decisions.

Table 1: Determinants of Drilling

	(1)	(2)	(3)
No. func. wells outside ref. plot ($\mathcal{N}$)	-0.0821	-0.533	-0.495
	(0.171)	(0.215)	(0.202)
1 func. well on ref. plot ($\mathbb{1}_{n=1}$)	-1.266	-1.310	-1.322
	(0.143)	(0.146)	(0.148)
2+ func. wells on ref. plot ($\mathbb{1}_{n=2}$)	-2.028	-2.159	-2.096
	(0.301)	(0.310)	(0.309)
Good monsoon (R)	-0.0536	-0.0368	-0.0477
	(0.179)	(0.180)	(0.191)
Log(ref plot area)	0.233	0.241	0.279
	(0.0772)	(0.0778)	(0.0823)
No. of plots in adjacency	-0.00658	-0.00951	0.0209
	(0.0585)	(0.0587)	(0.0594)
Mandal dummies	No	No	Yes
Estimation method	CRE	CRE-MEM	CRE-MEM
Log-likelihood	-1093	-1090	-1077
p-value (H_0 : No heterogeneity)	0.497	0.000	0.001

Notes: Standard errors in parentheses. Dependent variable is an indicator for whether well-drilling was attempted on reference plot that year (mean= 0.097). All estimations use a sample of 763 reference plots over five years, for a total sample of 3,815. All logit models estimated by maximum likelihood (selected coefficients reported). For estimation method, CRE refers to correlated (normally distributed) random effects (see Section 3.1) and MEM to misclassification error model (Section 3.2).

Table 1 also shows that having more functional borewells on the reference plot substantially reduces well-drilling by the plot owner, presumably due to diminishing returns to groundwater in production and to potential interference among own borewells.[19] A good past monsoon does not significantly affect the propensity to drill, as would be expected if rainfall were iid across years and thus not predictive of future rainfall. Drilling is more likely on larger plots, but does not depend on how many plots

[18] Misclassification error thus appears to act like classical measurement error leading to attenuation bias. Indeed, in Appendix F, we obtain qualitatively similar results from a linear probability model with plot fixed effects using instrumental variables to correct for classical measurement error.

[19] Predicted annual drilling rate fall from 0.212 to 0.057 to 0.026 in going from 0 to 1 to 2 reference plot borewells.

there are in the adjacency, which supports a simplified state-space $(\mathcal{N}, n)$ for the reference plot owner's investment decision (see Section 4.2 below). The inclusion of dummies for the 12 mandals (col. 3) has little effect on any of the results in Table 1.

For each specification, we perform a likelihood ratio test against the null of no unobserved heterogeneity at the reference plot level, referring specifically to the distribution of unconditional heterogeneity ν_i (see the next subsection for details of the testing procedure). In both CRE-MEM specifications, we strongly reject the null of no heterogeneity (bottom row of Table 1).[20] This finding suggests that unobserved heterogeneity explains part of the cross-sectional correlation between drilling and the number of functioning wells in the adjacency and should thus be addressed in the structural estimation.

3.4 A joint model of well flow and failure

The well flow and failure panels cover 382 and 697 adjacencies, respectively, of which 360 overlap, i.e., include wells with both flow and failure observations.[21] This overlap allows identification of the correlation in reference plot-level unobserved heterogeneity between the well flow and failure processes. Such correlation is plausible if well failure is seen as a state of zero flow forever.

First, we discuss the likelihood contribution of each process in turn and then derive the joint flow-failure likelihood.

Flow To estimate the probabilities for the five well-flow states ($q = 0.1, 0.25, 0.5, 0.75, 1.0$), we use a CRE ordered logit for the two-year panel. The conditional likelihood contribution of reference plot i is

$$\ell_i^f(\mu) = \prod_t \prod_{m=1}^{5} \left(\frac{1}{1 + e^{c_{m+1} + z_{it}^f(\mu)}} - \frac{1}{1 + e^{c_m + z_{it}^f(\mu)}} \right)^{\mathbb{1}_{Q_{it}=m}} \tag{9}$$

where $z_{it}^f(\mu)$ is a linear index for flow as in equation (4), Q_{it} is a 5-valued flow-state indicator and the c_m are cutoff parameters with $c_1 = -\infty$ and $c_6 = \infty$.[22]

[20]The p-values account for testing on the boundary of the parameter space; i.e., they are one-half of the probability that a chi-squared with 1 degree of freedom is greater than the LR test statistic.

[21]Non-overlap occurs because flow data were collected on all borewells owned by the household, irrespective of their inclusion in the adjacency survey, and because there are adjacencies that did not have functioning wells on the reference plot in 2010 and 2017, when flow data were collected.

[22]$\mathcal{N}_{it}$ differs between flow and failure estimation datasets. In the former, farmers were asked about the number of functioning borewells within a 100 meters radius of the reference plot, which differs from an adjacency, although the two definitions yield the same number of surrounding borewells on average. However, since the $\mathcal{N}_{it}$ used in the flow estimation is the contemporaneous (rather than retrospective) report of the respondent, we assume no misclassification error in (9).

Failure For reasons discussed in Section 2.4 (and Appendix E), we adopt a constant failure hazard specification, using the sequential logit as in Cameron and Heckman (1998), among others. The conditional likelihood contribution with $\mathcal{N}_{it}$ subject to misclassification error is

$$\ell_i^F(\xi) = \prod_{t=\tau_i}^{T_i} \sum_{k=1}^{\mathcal{N}_{it}^E} \frac{e^{z_{it}^F(k,\xi)\cdot F_{iat}}}{1+e^{z_{it}^F(k,\xi)}} \cdot Pr(k|\mathcal{N}_{it}^E, \hat{p}), \tag{10}$$

where $z_{it}^F(\xi)$ is a linear index for failure, F_{it} is a binary failure indicator, τ_i is the year that the borewell first enters the panel (or 2012, whichever comes last), T_i is the last year the borewell exists in the panel (or 2016, whichever comes first), and ξ is the unobserved heterogeneity in well failure. As mentioned in Section 2.2, for reference plots with multiple wells, only the first one sunk is included in the failure panel. Allowing multiple borewells on a plot would lead to a violation of strict exogeneity due to correlation between N_{it} and the failure shock.

The joint model For the joint flow/failure estimation, we follow, e.g., Eckstein and Wolpin (1999) in assuming that the reference plot level random effects, μ and ξ, are linearly related, i.e., $\xi = \kappa\mu$, where κ is a covariance parameter. Defining three indicator variables, D_i^1, D_i^2, D_i^3 for whether reference plot i contributes, respectively, only flow data, only failure data, or both flow and failure data, and assuming that μ is normally distributed with variance σ_μ^2, the full log-likelihood is

$$\mathcal{L} = \sum_i \log\left\{\int_\mu \ell_i(\mu,\kappa\mu)d\Phi(\frac{\mu}{\sigma_\mu})\right\}, \tag{11}$$

where

$$\ell_i(\mu,\xi) = \left[\ell_i^f(\mu)\right]^{D_i^1}\left[\ell_i^F(\xi)\right]^{D_i^2}\left[\ell_i^f(\mu)\ell_i^F(\xi)\right]^{D_i^3}. \tag{12}$$

We use 10-point Guass-Hermite quadrature to integrate out the continuous random effect μ.[23]

To estimate the probabilities of the five well flow states, $\pi_k(N,R,\nu^f)$, and the failure probability, $\pi_F(N,R,\nu^F)$, where ν^f and ν^F are, respectively, the flow and failure unobserved heterogeneity *unconditional* on the CRE covariates $(\bar{N}_i, \bar{R}_m)$, we proceed in four steps:

Step 1: Maximize the CRE likelihood given by equation (11) and obtain estimates of the linear index coefficients $\hat{\beta}^f, \hat{\gamma}^f, \hat{\beta}^F$, and $\hat{\gamma}^F$ (see equation 4).

[23]Compared to a discrete distribution, a continuous distribution of the random effect is easier to estimate and more conducive to hypothesis testing. However, since the structural model requires discrete types, we estimate a discrete heterogeneity distribution in our final specification (see Step 4 below).

Table 2: Joint Flow-Failure CRE Estimation

Step 1	(1)	(2)	(3)
Flow:			
log(N)	-0.849	-0.858	-0.884
	(0.171)	(0.172)	(0.172)
Good monsoon	1.766	1.782	1.917
	(0.800)	(0.802)	(0.804)
Failure:			
log(N)	0.052	1.842	1.163
	(0.635)	(0.560)	(0.473)
Good monsoon	0.119	0.127	-0.258
	(0.211)	(0.222)	(0.259)
Mandal dummies	NO	NO	YES
Estimation method	CRE	CRE-MEM	CRE-MEM
Log-likelihood	-2,246.71	-2,240.36	-2,163.23
Step 2			
σ_ν	1.289	1.390	0.311
	(0.093)	(0.112)	(0.136)
κ_ν	0.059	-1.864	-4.676
	(0.205)	(0.237)	(2.220)
ρ_ν	0.024	-0.498	-0.106
	(0.084)	(0.029)	(0.045)
Log-likelihood	-2404	-2481	-2247
p-value (H_0 : No het.)	0.000	0.000	0.000

Notes: Standard errors in parentheses. Maximum likelihood estimates with reference plot-level correlated random effects (CRE). Ordered logit cutoffs for flow, constant term for failure, and CRE covariate coefficients for both equations, not reported. Sample size = 3,401. σ_ν is standard deviation of (unconditional) unobserved heterogeneity; κ_ν is flow-failure covariance of same; $\rho_\nu = corr(\nu^f + \varepsilon^f, \nu^F + \varepsilon^F)$ is full cross-equation error correlation.

Step 2: Set $\beta^f = \hat{\beta}^f$, $\beta^F = \hat{\beta}^F$, $\gamma^f = \gamma^F = 0$, and re-maximize the likelihood with respect to the unconditional heterogeneity distribution parameters $\sigma_\nu = \sqrt{var(\nu^f)}$ and $\kappa_\nu = cov(\nu^f, \nu^F)/\sigma_\nu^2$.

Step 3: Test $H_0 : \sigma_\nu = \kappa_\nu = 0$.[24] If reject, go to Step 4. Otherwise, set $\nu^f = \nu^F = 0$.

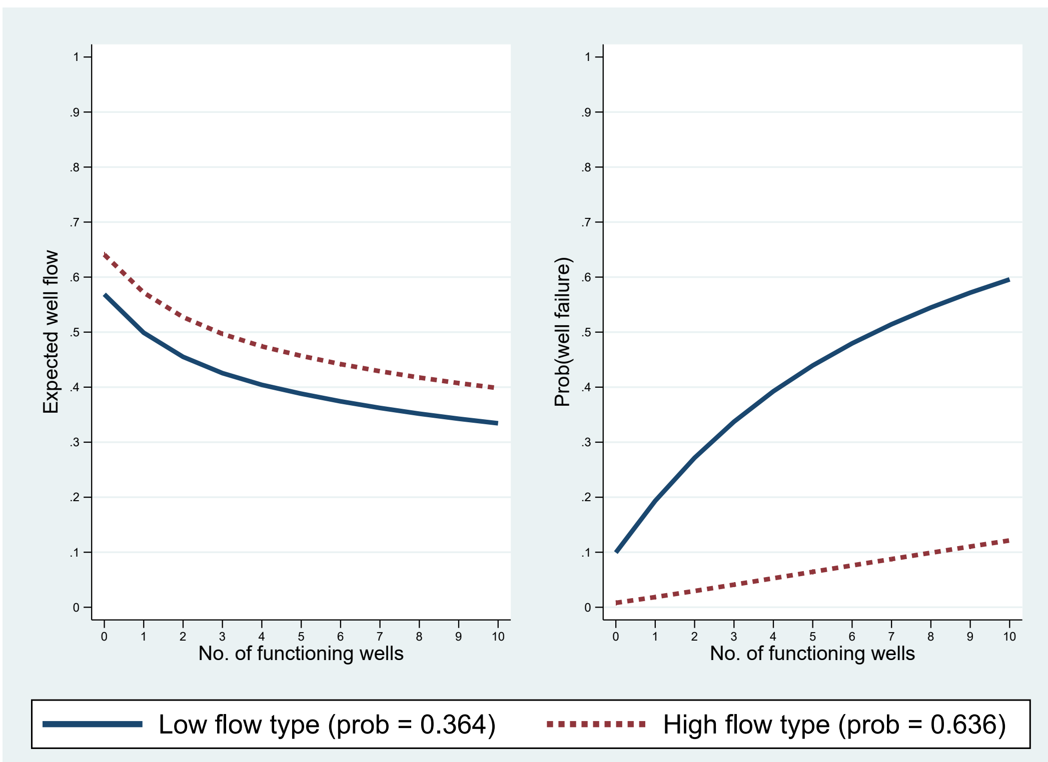
Step 4: Estimate a discrete joint distribution of (ν_f, ν_F) adding points of support $j = 1, ..., J$ until the

[24] This test presents the same boundary condition problem encountered earlier in Table 1 complicated further by κ_ν not being identified under the null. Following Stata's advice for such scenarios (see "help j_mixedlr" and citations therein), we use a conventional chi-square statistic to obtain a conservative p-value.

likelihood fails to improve. Compute $\pi_k(N, R, \nu_j^f)$ and $\pi_F(N, R, \nu_j^F)$ for each j.

The top panel of Table 2 reports the coefficient estimates from Step 1. Column 1 ignores misclassification error, column 2 corrects for misclassification error using the MEM approach, and column 3 adds mandal dummies to the column 2 specification. Corroborating the well interference externality, we find that having more borewells in an adjacency depresses flow and makes failure of the reference well more likely. This latter effect, however, only emerges with the MEM estimation in columns 2 and 3. Also, having had a good previous monsoon improves well flow but does not have a significant effect on failure, consistent with our interpretation of well failure as an absorbing state, independent of the vagaries of the monsoon. Including mandal dummies (in both flow and failure indices) shrinks the estimated scale of unobserved heterogeneity σ_ν from 1.39 to 0.31. Lastly, only the MEM specifications in columns 2 and 3 show the expected negative correlation between flow and failure heterogeneity and, in both specifications, we strongly reject the null of no unobserved heterogeneity (Step 3).

Figure 4: Expected flow and failure probability by type



Notes: The left panel displays by unobserved type the expected well flow $\sum_k \pi_k q_k$ as a function of N (number of functioning wells in adjacency) as predicted from the joint flow/failure model. The right panel shows, also by unobserved type, the predicted failure probability as a function of N. Probability functions are averaged over mandals and rainfall states.

Moving to Step 4, we redo Step 2 allowing for 2 *discrete* types, obtaining a log-likelihood value of -2245.59 (compared to -2246.55 in column 3 of Table 2). Since adding a third type does not lead to an appreciable improvement in the likelihood, we stop at $J = 2$ and compute the flow and failure probabilities. For each unobserved flow type, "low" and "high", Figure 4 plots expected well flow, $\sum_k \pi_k q_k$, on the left panel and the probability of well failure, π_F, on the right panel against the (hypothetical) number of functioning wells in the adjacency N, averaging across mandals and rainfall states for ease of presentation. While expected flow differs modestly between high and low unobserved flow types, the *marginal* effect of N on expected flow (the intensive margin externality) is virtually identical across types. By contrast, both the rate of well failure and the marginal effect of N on failure (extensive margin externality) are higher for the low flow type (probability = 0.364) than for the high flow type (probability = 0.636).

To summarize the empirical results thus far, having more functional borewells adjacent to a reference plot reduces the discharge of borewells on that plot, increase their likelihood of failure, and reduces the propensity for further drilling. Taken together, these findings point to a well interference externality that farmers account for in their investment decisions.

4 A Model of Borewell Investment

In this section, we present a dynamic equilibrium model of borewell investment on a large plot network; Section 5 will describe the structural estimation of its parameters. For reasons discussed in Section 2, we assume that our model applies only to developed plots and that the owners of such plots do not face liquidity constraints when deciding each year whether or not to drill.

4.1 Preliminaries

Let the *incremental* agricultural output generated from a functioning borewell on plot i at time t be

$$y_{it} = \theta \left[\alpha q_{it}^{\delta} + (1-\alpha) a_i^{\delta} \right]^{\frac{1}{\delta}} , \tag{13}$$

where (θ, α, δ) are parameters, a_i is plot area, and q_{it} is well discharge or flow. Flow is stochastic and thus unknown to the farmer prior to drilling and has a discrete distribution with K points of support $\{q_{it1}, ..., q_{itK}\}$ each with probability π_{itk}. Along with constant elasticity of substitution, production function (13) imposes constant returns to scale (CRS); i.e., output per acre depends only on flow per acre.[25] The scale parameter θ converts physical output into 2017 Indian rupees (Rs).

Year t flow state probabilities π_{itk} depend on past monsoon rainfall (i.e., aquifer recharge) and on the number of functioning borewells in the adjacency at the beginning of year t according to

$$\pi_{itk} = \pi_k(N_{it}, R_{t-1}). \tag{14}$$

The well interference externality, estimated in Section 3.4, can be thought of as a higher N_{it} shifting the probability mass to low flow states. A borewell remains functional, with positive discharge, until stochastic failure occurs with probability $\pi_{Fit} = \pi_F(N_{it}, R_{t-1})$. Since failure is an absorbing state, the failed borewell henceforth contributes nothing further to output.[26]

Drilling success is also stochastic and, for reasons discussed in Section 2.6, its probability π_S is assumed constant. Each drilling attempt entails a fixed cost c_d for sinking the bore hole and, if the attempt is successful, there is an additional cost of installing a pipe, casing, and hooking up the electrical connection; the submersible pump itself is removable and thus is not considered a sunk cost. The total cost of a successful attempt is, therefore, $c_s > c_d$.

Finally, for the sake of tractability, we assume that at most two wells can function simultaneously on any given plot so that $N_{it} \in \{0, .., 2p_i\}$, where p_i is the number of plots in adjacency i.[27] Drilling success, failure, and discharge events for two wells on the same plot are independent random variables (*conditional* on the plot-specific unobserved heterogeneity described in Section 3.4). Using superscripts to enumerate wells, incremental output of a plot with two wells depends on the sum of their discharges $q^1_{it} + q^2_{it}$, since water from both wells can be pooled and dispersed throughout the plot.

Summarizing, expected output conditional on monsoon rainfall may be written as

$$\begin{aligned}
\mathbb{E}[y_{it}(N_{it}, n_{it})|R_{t-1}] &= \sum_{k=1}^{K} \pi_{itk}(N_{it}, R_{t-1})\theta \left[\alpha (q^1_{itk})^{\delta} + (1-\alpha) a_i^{\delta}\right]^{\frac{1}{\delta}} \qquad \text{if } n_{it} = 1 \\
&= \sum_{j=1}^{K}\sum_{k=1}^{K} \pi_{itj}(N_{it}, R_{t-1})\pi_{itk}(N_{it}, R_{t-1})\theta \left[\alpha (q^1_{itk} + q^2_{itj})^{\delta} + (1-\alpha) a_i^{\delta}\right]^{\frac{1}{\delta}} \qquad \text{if } n_{it} = 2.
\end{aligned} \tag{15}$$

[25] The Online Appendix of Giné and Jacoby (2020) tests and cannot reject CRS based on a Cobb-Douglas production function estimation in a closely related setting.

[26] While we allow the failure probability to depend on rainfall from the past monsoon for the sake of generality, a null effect of rainfall is more consistent with well failure being an absorbing state, which is indeed what we find in Section 3.4.

[27] In our representative plot panel, 3 or more functioning wells occurs in just 35 out of all 14,310 plot-years.

4.2 Borewell investment decision

We now consider the discrete choice to drill ($d = 1$) or not to drill ($d = 0$) and derive the plot owner's decision rule or conditional choice probability $CCP(\mathcal{N}, n) \equiv \Pr(d = 1 \mid \mathcal{N}, n)$, temporarily dropping subscripts for ease of exposition. We first describe the dynamic decision facing the owner of a plot with area a in an adjacency with p plots, in isolation, i.e., taking as given their beliefs about the evolution of the state of the adjacency. As noted, the state space of the plot owner consists only of the total number of wells in the other plots of the adjacency $\mathcal{N} \in \{0, .., 2(p-1)\}$ and the number of own functioning wells, $n \in \{0, 1, 2\}$. In the next subsection and later in Subsection 5.1, we discuss this assumption and its role in a tractable equilibrium model of beliefs and conditional choice probabilities.

By assumption, state $n = 0$ or $n = 1$ are the only cases where investment can occur. A plot owner with $n = 0$ may decide not to drill, with payoff value $\overline{v}_{00}(\mathcal{N}) + \epsilon_{00}$, or to drill, with payoff value $\overline{v}_{0I}(\mathcal{N}) + \epsilon_{0I}$. As in a random-utility framework, choice-specific payoffs have additive "deterministic" and "random" components. The random components of the payoff of waiting (ϵ_{00}) or drilling (ϵ_{0I}) are realized every period before choices are made, iid across choices and time, and unobserved by other plot owners in the adjacency, each of whom are drawing their own random components.

The deterministic components, which are known to the plot owner conditional on the observable state variables and parameters, include the static one-period profits (expected value of output minus drilling costs, if any) and the expected continuation values. For the no drilling (waiting) choice,

$$\begin{aligned} \overline{v}_{00}(\mathcal{N}) &= \beta \, \mathbb{E} \, V(\mathcal{N}', 0) \\ &= \beta \sum_{\mathcal{N}'} \tilde{F}(\mathcal{N}' \mid \mathcal{N}, 0) V(\mathcal{N}', 0) \end{aligned} \tag{16}$$

and for the choice of making a drilling attempt

$$\begin{aligned} \overline{v}_{0I}(\mathcal{N}) = \pi_S &\left(-c_s + \beta \sum_{\mathcal{N}'} \tilde{F}(\mathcal{N}' \mid \mathcal{N}, 0) V(\mathcal{N}', 1) \right) \\ &+ (1 - \pi_S) \left(-c_d + \beta \sum_{\mathcal{N}'} \tilde{F}(\mathcal{N}' \mid \mathcal{N}, 0) V(\mathcal{N}', 0) \right), \end{aligned} \tag{17}$$

where the value function $V(\mathcal{N}, n)$ is defined below, β is the discount factor and $\tilde{F}(\mathcal{N}' \mid \mathcal{N}, n)$ reflects beliefs about the probability of $\mathcal{N}'$ functioning wells in other adjacency plots next period conditional on $\mathcal{N}$ functioning wells in other adjacency plots and on n functioning wells on the reference plot ($n = 0$, in this case) this period. Since drilling occurs after rabi season production (see Figure 2), the increase in expected output from any successful attempt is only realized in the next period.

We assume that the random components associated with the choices of waiting and drilling, $(\epsilon_{00}, \epsilon_{0I})$, are each iid Type-1 extreme value with location parameter 0 and scale parameter σ. Further, denote by $V(\mathcal{N}, n)$ the beginning-of-period value function for the plot owner, before these random components of payoffs are realized. Taking expectations for $n = 0$, we have

$$\begin{aligned} V(\mathcal{N}, 0) &= \mathbb{E} \max \Big\{ \overline{v}_{00}(\mathcal{N}) + \epsilon_{00}, \overline{v}_{0I}(\mathcal{N}) + \epsilon_{0I} \Big\} \\ &= \sigma \Big(\gamma + \log \big(\exp(\overline{v}_{00}(\mathcal{N})/\sigma) + \exp(\overline{v}_{0I}(\mathcal{N})/\sigma) \big) \Big) \end{aligned} \tag{18}$$

where the second line follows from the Type-1 extreme value assumption and γ is Euler's constant.

Similarly, a borewell owner with $n = 1$ may decide to wait, with payoff value $\overline{v}_{10}(\mathcal{N}) + \epsilon_{10}$, where

$$\begin{aligned} \overline{v}_{10}(\mathcal{N}) = \mathbb{E} \Bigg\{ & \mathbb{E}[y(\mathcal{N}+1, 1)|R] \\ & + \beta \Bigg((1 - \pi_F(\mathcal{N}+1, R)) \sum_{\mathcal{N}'} \tilde{F}(\mathcal{N}' \mid \mathcal{N}, 1) V(\mathcal{N}', 1) \\ & + \pi_F(\mathcal{N}+1, R) \sum_{\mathcal{N}'} \tilde{F}(\mathcal{N}' \mid \mathcal{N}, 1) V(\mathcal{N}', 0) \Bigg) \Bigg\}, \end{aligned} \tag{19}$$

using equation (15) for the inner expectation of output conditional on monsoon rainfall R and taking the outer expectation with respect to the distribution of R. Alternatively, the plot owner may attempt to drill a second borewell, with payoff value $\overline{v}_{1I}(\mathcal{N}) + \epsilon_{1I}$, where

$$\begin{aligned} \overline{v}_{1I}(\mathcal{N}) = \mathbb{E} \Bigg\{ & \mathbb{E}[y(\mathcal{N}+1, 1)|R] - c_s \pi_S - c_d (1 - \pi_S) \\ & + \beta \Bigg(\pi_S (1 - \pi_F(\mathcal{N}+1, R)) \sum_{\mathcal{N}'} \tilde{F}(\mathcal{N}' \mid \mathcal{N}, 1) V(\mathcal{N}', 2) \Bigg) \\ & + \beta \left(\pi_S \pi_F(\mathcal{N}+1, R) + (1 - \pi_S)(1 - \pi_F(\mathcal{N}+1, R)) \right) \\ & \qquad \times \sum_{\mathcal{N}'} \tilde{F}(\mathcal{N}' \mid \mathcal{N}, 1) V(\mathcal{N}', 1) \\ & + \beta (1 - \pi_S) \pi_F(\mathcal{N}+1, R) \sum_{\mathcal{N}'} F(\mathcal{N}' \mid \mathcal{N}, 1) V(\mathcal{N}', 0) \Bigg\}. \end{aligned} \tag{20}$$

We can now write

$$
\begin{aligned}
V(\mathcal{N},1) &= \mathbb{E}\max\Big\{\overline{v}_{10}(\mathcal{N})+\epsilon_{10},\overline{v}_{1I}(\mathcal{N})+\epsilon_{1I}\Big\} \\
&= \sigma\big(\gamma+\log\left(\exp(\overline{v}_{10}(\mathcal{N})/\sigma)+\exp(\overline{v}_{1I}(\mathcal{N})/\sigma)\right)\big)
\end{aligned} \tag{21}
$$

where the second line follows, again, from an analogous Type-1 extreme value assumption on $(\epsilon_{10},\epsilon_{1I})$.

Finally, we have

$$
\begin{aligned}
&V(\mathcal{N},2)=\mathbb{E}\Big\{\,\mathbb{E}[y(\mathcal{N}+2,2)|R] \\
&+\beta\Bigg((1-\pi_F(\mathcal{N}+2,R))^2\sum_{\mathcal{N}'}\tilde{F}(\mathcal{N}'\mid\mathcal{N},2)V(\mathcal{N}',2) \\
&+2\pi_F(\mathcal{N}+2,R)(1-\pi_F(\mathcal{N}+2,R))\sum_{\mathcal{N}'}\tilde{F}(\mathcal{N}'\mid\mathcal{N},2)V(\mathcal{N}',1) \\
&+\pi_F^2(\mathcal{N}+2,R)\sum_{\mathcal{N}'}\tilde{F}(\mathcal{N}'\mid\mathcal{N},2)V(\mathcal{N}',0)\Bigg)\Big\}.
\end{aligned} \tag{22}
$$

Note that equations (16)-(22) combine to form the Bellman equation for this investment problem.

The discrete choice to attempt drilling a borewell in the reference plot is thus

$$
d=d(\mathcal{N},n)=\begin{cases} 1 & \text{if } n<2 \text{ and } \overline{v}_{nI}(\mathcal{N})-\overline{v}_{n0}(\mathcal{N})>\epsilon_{n0}-\epsilon_{nI} \\ 0 & \text{otherwise,}\end{cases}
$$

using equations (16), (17), (19) and (20). With logit random utility shocks, the decision rule as perceived by the researcher (and by neighbors) is characterized by the CCP function

$$
\begin{aligned}
\mathrm{CCP}(\mathcal{N},n) &= \Pr(d=1\mid\mathcal{N},n)=\Pr(\epsilon_{n0}-\epsilon_{nI}<\overline{v}_{nI}(\mathcal{N})-\overline{v}_{n0}(\mathcal{N})) \\
&= \frac{\exp(\overline{v}_{nI}(\mathcal{N})/\sigma)}{\exp(\overline{v}_{nI}(\mathcal{N})/\sigma)+\exp(\overline{v}_{n0}(\mathcal{N})/\sigma)}.
\end{aligned}
$$

4.3 Adjacency equilibrium

Before characterizing the equilibrium of the dynamic drilling game, we introduce the concept of a village "map", or plot network, upon which this game is played. We use 14 cadastral maps representing at least one village in each mandal (see Appendix D). While the borders of these administrative maps are arbitrary in that they do not correspond to salient geographic or geological features, each contains many plots and as a result "truncation-at-border" effects should have negligible empirical consequences.[28]

Formally, a cadastral map with P plots is characterized by a $P\times 1$ vector A listing the area of each plot and a $P\times P$ adjacency matrix $\mathbf{M}$ with typical element $M_{ij}=1$ if plot j adjoins plot i and 0 otherwise, and with $M_{ii}=1$. Ignoring for now heterogeneity in the form of the plot's developed status and flow/failure type, $\{A,\mathbf{M}\}$ fully characterizes all adjacencies in the map. For instance, plot i has an area equal to the i-th element of A and its adjacency has $\sum_j M_{ij}$ plots because plot j with $M_{ij}=1$ belongs in plot i's adjacency. Let $\mathcal{M}_{(ih)}$ be the set of plots h-level adjacent to plot i so that $\mathcal{M}_{(i1)}=\{j: M_{ij}=1\}$ is the set of immediate (1-level) neighbors in i's adjacency, $\mathcal{M}_{(i2)}=\left\{j: j\notin\mathcal{M}_{(i1)}, j\in\mathcal{M}_{(k1)}, k\in\mathcal{M}_{(i1)}\right\}$ is the set of 1-level adjacent neighbors of i's 1-level adjacent neighbors, and so on for all "layers" h.

Let the state of plot i in period t be the number of functioning wells on the plot at the beginning of the year $n_{it}\in\{0,1,2\}$. Further, let $X_t=\{n_{it}: i=1,\ldots,P\}$ be the state of the map, representing the entire spatial distribution of borewells in the cadastral map. Now, define $X_{(ih)t}=\left\{n_{jt}: j\in\mathcal{M}_{(ih)}\right\}$, where $X_{(i1)t}$ collects the state of the neighbors of reference plot i, $X_{(i2)t}$ collects the state of the neighbors' neighbors, and so on.

Thus far, we have taken beliefs about the evolution of the number of functioning wells in the adjacency as given, deriving the plot owner's dynamic investment decision as if it were a game "against nature". However, we assume a Markov-perfect equilibrium (MPE), in which beliefs and decision rules (CCPs) of all plot owners are consistent with one another. Furthermore, our state space $(\mathcal{N},n)$ implicitly assumes that plot owners ignore the status of wells on successive layers of plots outside their own adjacencies. This restriction is not, in general, implied by our key assumption that well interference is limited to functioning wells in the adjacency. Indeed, information on the status of wells outside the first layer might help agents predict neighbors' investment behavior and the status of wells in their adjacencies, which in turn helps neighbors predict their neighbors' investment behavior and the status of their wells, and so on. Under *unrestricted* MPE play, therefore, investment decisions may depend on the state of the whole map, even with well interference effects confined to adjacent plots. To be precise, let $CCP_i(X_t)$ be a choice probability function for the owner of plot i and $\{CCP\}$ be the vector of choice probabilities of all plot owners in the cadastral map. Further, let one-period ahead transition probabilities $\widetilde{F}(X_{t+1}\mid X_t)$ describe beliefs about the evolution of the state of the map and $F(X_{t+1}\mid X_t;\{CCP\})$ be the one-period-ahead law of motion for the state induced by the primitives of the problem and $\{CCP\}$. We thus have:

Definition 1. An MPE is a vector of choice probabilities $\{CCP_i^*(X_t): i=1,\ldots,P\}$ and beliefs $\widetilde{F}^*()$ such that: a) given beliefs $\widetilde{F}^*()$, $CCP_i^*()$ is the solution of plot owner i's dynamic game "against nature";

[28]To be sure, adjacencies of border plots will always be truncated. However, our average cadastral map has 881 plots with only 102 (12%) being border plots.

and b) beliefs are correct, in that $\widetilde{F}^*(X_{t+1} \mid X_t) = F(X_{t+1} \mid X_t; \{CCP\}^*)$.

In general, each plot owner in their unique adjacency would have a different equilibrium CCP depending on all primitives, including the structure of the map. Given the number of plots in the map, unrestricted MPE play is not empirically feasible due to the high dimensionality of $\{X_t, \{CCP\}\}$.

As a tractable alternative, we consider a Markov equilibrium in which: i) CCPs depend only on the state of the (1-level) adjacency $(X_{(i1)t}, n_{it})$, and ii) the plot owner has beliefs only about the stochastic evolution of $X_{(i1)t}$ in steady state. While assumption i) avoids the "curse of dimensionality", the fact that well interference is largely limited to the adjacency should dampen the influences induced by unrestricted play of plot owners in layers $h > 1$ as well as making it less plausible (i.e., by bounded rationality) that plot owners would keep track of the full state of a large map. Assumption ii) is not only a natural implication of assumption i), but it also adds the non-trivial requirement that equilibrium beliefs about the state of the adjacency be correct when averaged over the map's stochastic steady state. Thus, in the spirit of an "oblivious equilibrium",[29] we propose

Definition 2. An Adjacency Equilibrium (AE) is a vector of choice probabilities $\{CCP_i^*(X_{(i1)t}, n_{it}) : i = 1, \ldots, P\}$ and of beliefs $\left\{\widetilde{F}_i^*(X_{(i1)t+1} \mid X_{(i1)t}, n_{it}) : i = 1, \ldots, P\right\}$ such that: a) given beliefs $\widetilde{F}_i^*()$, the decision rule $CCP_i^*()$ is the solution of plot owner i's dynamic game "against nature"; and b) beliefs are correct "on average" in steady state. That is, let $F^\infty(X_t; \{CCP\})$ be the stationary joint distribution over the state induced by the primitives and the vector of CCPs.[30] Further, let $F_i^\infty(X_{(i2)t} \mid X_{(i1)t}, n_{it}; \{CCP\})$ be the conditional distribution implied by $F^\infty(X_t; \{CCP\})$. Then,

$$\begin{aligned} &\widetilde{F}_i^*(X_{(i1)t+1} = x_{(i1)t+1} \mid X_{(i1)t} = x_{(i1)t}, n_{it}) = \\ &\sum_{x_{(i2)t}} F_i^\infty(x_{(i2)t} \mid x_{(i1)t}, n_{it}; \{CCP\}^*) F_i(x_{(i1)t+1} \mid x_{(i1)t}, x_{(i2)t}, n_{it}; \{CCP\}^*). \end{aligned} \tag{23}$$

To understand how equation (23) constrains beliefs, note first that the evolution of the state of plot $j \in \mathcal{M}_{(i1)}$ between t and $t+1$ depends on $CCP_j^*()$ at t. This investment decision rule depends, in turn, upon the state of j's adjacency at t, formed by plot j and all of its neighbors, including plot i. All of the plots in j's adjacency are in $\mathcal{M}_{(i1)}$ and $\mathcal{M}_{(i2)}$. Therefore, the state variables of plot owner j are contained in $\{n_{it}, X_{(i1)t}, X_{(i2)t}\}$. If the owner of plot i knew $X_{(i2)t}$, they would thus be able to predict their neighbor j's behavior at t using $CCP_j^*()$ and, together with other primitives such as the

[29]Weintraub et al. (2008), Benkard et al. (2015) and Ifrach and Weintraub (2017) consider alternative "oblivious equilibrium" concepts in the context of the Ericson and Pakes (1995) model of industry dynamics and show that they closely approximate the corresponding MPE. While we expect similar approximation results to hold in our setting, we leave this issue for future research.

[30]Since the state of the map is an irreducible and aperiodic Markov chain, a unique stationary distribution exists.

drilling success and well failure processes, predict the stochastic evolution of the state of plot j which is the second factor in each term of the summation in equation (23). An AE, however, assumes that $X_{(i2)t}$ in not in plot owner i's information set, but rather that they can form expectations about it using the (conditional) steady state distribution $F^{\infty}(X_{(i2)t} \mid X_{(i1)t}, n_{it}; \{CCP\}^*)$ as probability weights on the RHS of equation (23).[31] Although each plot owner still has a unique CCP and set of beliefs, and the plot owners' joint decisions still depend on the entire cadastral map, the AE concept achieves considerable simplification.[32]

5 Structural Estimation

We now detail a tractable empirical structural model and a Simulated Method of Moments (SMM) procedure to estimate it. Three earlier results inform our estimation strategy: First, since no drilling activity is observed on most plots, it is reasonable to assume that the model of optimal annual drilling choice outlined in Section 4 is only relevant to the 38.7% of plots in the representative sample that are developed (Section 2.2). Second, misclassification error in the reported number of functioning borewells on neighboring plots leads to non-trivial econometric biases (Section 3.3), greatly complicating a likelihood-based estimator. Third, our estimation algorithm has to account for the important plot-level unobserved heterogeneity in groundwater availability driving both well flow and failure (Section 3.4). In addition to the four structural parameters $\Omega = (\theta, \alpha, \delta, \sigma)$ that we recover via SMM, Table 3 summarizes the model primitives derived from the first stage estimation of Subsection 3.4 (i.e., the joint flow/fail model) and from other sources.

Discount factor Given the challenge of identifying the discount factor in dynamic discrete choice models (Rust 1994; Magnac and Thesmar 2002), we follow standard practice by fixing the value of β in the SMM estimation. However, since our welfare calculations may be sensitive to this choice, we calibrate β: First, we estimate the structural model at different (fixed) values of β along a coarse grid. Next, we simulate from each of these estimated models in steady state an average of the present discounted value of developed land per acre, noting that the value of undeveloped land is normalized to zero. Lastly, we compare this value differential between developed and undeveloped land to its empirical

[31] In our empirical implementation, we do not use equation (23) directly but rather compute equilibrium beliefs using "brute force" by simulating very long histories of investment, success, and failure events for every plot in each cadastral map until a steady state is reached. We then use simulated histories to compute the requisite transition probabilities.

[32] Using Brouwer's fixed point theorem, we can show that at least one AE exists. Multiplicity of equilibria, however, cannot be ruled out. Xu (2018) establishes that, in a static version of a similar model, the best response operator has a contraction property provided that the "strategic interaction parameter" is small enough. An extension of this result to a dynamic setting is nontrivial and is left as a topic for future research.

Table 3: MODEL PRIMITIVES

	Symbol(s)	Subsection/note
Estimated in 2nd stage:		
Production function	θ, α, δ	4.1
Scale of drilling shock	σ	4.2
Estimated in 1st stage:		
Flow state probability functions	$\pi_1, ..., \pi_5$	3.4/note 1
Failure probability function	π_F	3.4/note 1
Flow/fail heterogeneity	ν_1, ν_2	3.4/note 2
Successful drilling cost	c_s	note 3
Unsuccessful drilling cost	c_d	note 4
$P(\text{good monsoon}\|\text{mandal})$	$\mathbb{E}\, R_{mt-1}$	3.1
$P(\text{Success}\|\mathcal{D}=1, \text{map-village})$	π_S	2.6
$P(\mathcal{D}=1\|a_k, \text{map-village})$		2.2/note 5
Fixed parameters:		
Discount factor	β	note 6
Plot network (by map-village)	$\{A, \mathbf{M}\}$	4.3

Notes: See subsection specified in column 3 and/or notes as follows: (1) Probability functions depend upon the number of functioning borewells in the adjacency and vary at the mandal-level, as well as by monsoon rainfall and unobserved type; (2) Probability of (low) type 1 = 0.364; (3) $c_d = 28,800$ Rs. is computed as median drilling cost (in 2017 Rs) across all borewells sunk since the year 2000; (4) $c_s = 59,800$ Rs. is computed as the median of the sum of drilling, pipe, casing, and electrical connection costs across all borewells sunk since the year 2000; (5) Estimated from a logit regression of developed plot indicator $\mathcal{D}$ on plot area quintile a_k and map-village dummies (see Appendix Table D.1); (6) Calibrated from land value data (see main text and Appendix G for details).

counterpart, which we estimate to be around 80,000 Rs./acre (see Appendix G for details), and select the β yielding the closest match. This procedure delivers a value of $\beta = 0.95$.

State space restrictions and plot/adjacency types To make the empirical model more tractable, we assume that the CCP in the AE depends on the area of the reference plot a and on the number, but not on the areas of adjacent plots. This restriction effectively reduces $X_{(i1)t}$ to $\mathcal{N}_{it} = \sum_{j \in \mathcal{M}_{(i1)t}} n_{jt}$, yielding state space $(\mathcal{N}_{it}, n_{it})$.[33] Given the one-to-one mapping between CCPs, beliefs, and reference plot (or, equivalently, adjacency) "types", tractability concerns also dictate limiting the number L of

[33] We show in Section 3.3 that, conditional on $(\mathcal{N}_{it}, n_{it})$, drilling decisions do not depend on the number of plots in the adjacency, which suggests that using the more fine-grained state space $(X_{(i1)t}, n_{it})$ would not improve model fit.

such types. By types of developed plots, we refer to heterogeneity in characteristics that are both observed (area, number of adjacent plots) and unobserved (low or high flow/failure heterogeneity) to the econometrician. Our discretization of reference plot area into quintiles coupled with the number of adjacent plots in the maps ranging from 1 to 7, yields 35 possible observed types, along with 2 unobserved types for a total of $L = 70$ types. Implicit in our specification of types is that plot owners observe neither the developed status nor the flow/failure type of the *other* plots in their adjacencies. In line with this, we assume that developed status and flow/failure type are iid across plots.[34]

5.1 Solution algorithm

Given values of parameters Ω, we obtain an AE on each of the 14 cadastral maps as follows:

Initialize the maps:

Step 1 Draw a $\mathcal{D}_j$ for each plot j from the binomial distribution with $P(\mathcal{D} = 1|a_k, \text{map-village})$ from the first stage (see Table 3).

Step 2 Assign each plot with $\mathcal{D}_j = 1$ an unobserved flow type ν_1 or ν_2, drawing from a binomial distribution with probability of (low) type $1 = 0.364$.[35]

Step 3 Assign each plot an initial number (zero) of functioning borewells $\{n_{j0} : j = 1, \ldots, P\}$ and an initial choice probability function (constant equal to 0.5) to each type $\{CCP_{l,0} : l = 1, \ldots, L\}$.

Iterate on beliefs and CCPs:

Step 4 Given $\{CCP_{l,q-1} : l = 1, \ldots, L\}$ at iteration $q = 1, 2, ...$, simulate the time-series of well drilling decisions, successes and (unobserved type-specific) failures in every plot on the map until the steady state is reached. Simulate $T = 150,000$ periods forward *in* steady state.

Step 5 From the steady state simulations, construct estimates of the one-period ahead state transition matrices $F(\mathcal{N}'|\mathcal{N}, n)$ for each type, averaging across plots on the map of the same type. Denote these estimates by $\widehat{F}_{lq}$.

[34] While positive spatial correlation in developed status and flow/failure type is plausible, incorporating it is unlikely to appreciably improve the fit of the structural model. Given this, along with the limited information about spatial correlation in our data, as well as the additional complexity involved, we leave this refinement for future work.

[35] Type probabilities recovered in the first stage flow/fail estimation should accurately reflect the distribution of types in the population of developed plots since only 8% of relevant developed plots are excluded from the flow/fail estimation sample (based on the various selection criteria discussed in Subsections 2.4 and 2.5). Also, to ensure a unique AE despite the inherent randomness of a particular map draw in Steps 1 and 2, we repeat these two steps ten times for each plot and pool the resulting data in computing beliefs in Steps 4 and 5.

Step 6 Given beliefs $\widehat{F}_{lq}$ and primitives, use policy iteration to compute new CCP's which solve the plot owner's game "against nature". Upon convergence of policy iterations, obtain a $\{CCP_{lq}\}$ satisfying the fixed point condition $CCP_{lq} = \Psi(CCP_{lk-1}, \widehat{F}_{lk}, \Omega)$ for all types, where Ψ is a policy iteration operator.

Convergence:

Step 7 If $\|CCP_q - CCP_{q-1}\|$ is small enough, done. If not, update q and return to Step 5. If CCPs converge, so do beliefs, which are a continuous function of CCPs.

Steps 1-7 are nested within a routine for minimizing the SMM criterion function with respect to Ω using a downhill simplex method.

5.2 Moment conditions and identification

We match observed annual drilling rates in the representative sample of plots by plot area quintile a_k and by the number of functioning borewells $n = 0, 1$ – empirical moments shown in panel (a) of Figure 3 – to their model-based counterparts.[36] Since, by assumption, no investment occurs once a plot has two borewells, we do not match drilling rates conditional on $n \geq 2$, resulting in a total of ten moments.[37] Although we do not target moments involving the number of functioning borewells outside of the reference plot (i.e., average drilling rates conditional on different values of $\mathcal{N}$), later we exploit the correlation between drilling and $\mathcal{N}$ in a model validation exercise.

Because we estimate all parameters associated with well interference externalities in the first stage, identification of the remaining parameters Ω is straightforward.[38] Heuristically, one may think about identification in terms of a static model wherein drilling decisions are made once and for all and there is no borewell failure. In this case, we have $P_{n,k} \equiv Pr(d = 1|n, a_k) = \text{logit}^{-1}(\{\theta[f(n+1, a_k; \alpha, \delta) - f(n, a_k; \alpha, \delta)] - \mathbb{E}\, c\}/\sigma)$, where $\mathbb{E}\, c = c_s\pi_s + c_d(1 - \pi_s)$ is the expected cost of drilling and $f = \frac{1}{\theta}\, \mathbb{E}[y(N, R, n, a_k; \alpha, \delta)]$ is expected (physical) output on a plot of area a_k with n functional borewells, where expectations are taken with respect to flow outcomes, beliefs about the number of functioning borewells in other adjacency plots, and monsoon rainfall. Given α and δ, the difference in drilling rates across any two (n, k) pairs with different expected physical outputs identifies

[36] Model-based drilling rates are averages across the 14 map villages weighted by the proportion of total plot area in the representative plot sample contributed by sample plots associated with that village. Matching drilling rates on developed plots (Figure 3, panel b) would lead to similar results since, by construction, model village maps are endowed with developed plots in the same proportion as in the representative plot sample.

[37] The SMM criterion function uses a diagonal weighting matrix consisting of the inverse moment variances.

[38] In other words, our second-stage estimation does not require (as in the canonical case considered by Bajari et al. 2015) identifying the effect of well-owner's neighbors' actions on well-owner's payoffs.

the ratio θ/σ.[39] Since $\mathbb{E}\,c$ is a known constant, we can then back out σ, and hence θ, from the average drilling rate at any (n, k). Given $\theta(\alpha,\delta)$ and $\sigma(\alpha,\delta)$, the remaining eight moment conditions yield more than enough equations to solve for α and δ. Intuitively, fixing a_k, differences in drilling rates at $n = 1$ and $n = 0$ capture diminishing returns to flow because, in log odds form, $\log \frac{P_{1,k}(1-P_{0,k})}{P_{0,k}(1-P_{1,k})} = \frac{\theta(\alpha,\delta)}{\sigma(\alpha,\delta)}\{[f(2, a_k; \alpha, \delta) - f(1, a_k; \alpha, \delta)] - [f(1, a_k; \alpha, \delta) - 0]\}$. Likewise, now fixing n, differences in drilling rates across area quintiles capture how the marginal product of flow varies with area because $\log \frac{P_{n,k}(1-P_{n,k'})}{P_{n,k'}(1-P_{n,k})} = \frac{\theta(\alpha,\delta)}{\sigma(\alpha,\delta)}\{[f(n+1, a_k; \alpha, \delta) - f(n, a_k; \alpha, \delta)] - [f(n+1, a_{k'}; \alpha, \delta) - f(n, a_{k'}; \alpha, \delta)]\}$.

5.3 Results and model validation

Table 4 reports the model parameter estimates along with their bootstrapped standard errors based on 100 replications of our estimation procedure.[40] We can reject a Cobb-Douglas production function (i.e., $\delta = 0$) with a very high degree of confidence. Figure 5 also shows that our model matches the targeted moments reasonably well.

Table 4: Structural Parameter Estimates

θ	α	δ	σ
17.61	0.65	0.62	1.07
(0.46)	(0.02)	(0.03)	(0.04)

Notes: Bootstrapped standard errors in parentheses. See equation (13) for definition of production function parameters (θ, α, δ); σ is scale of drilling shock.

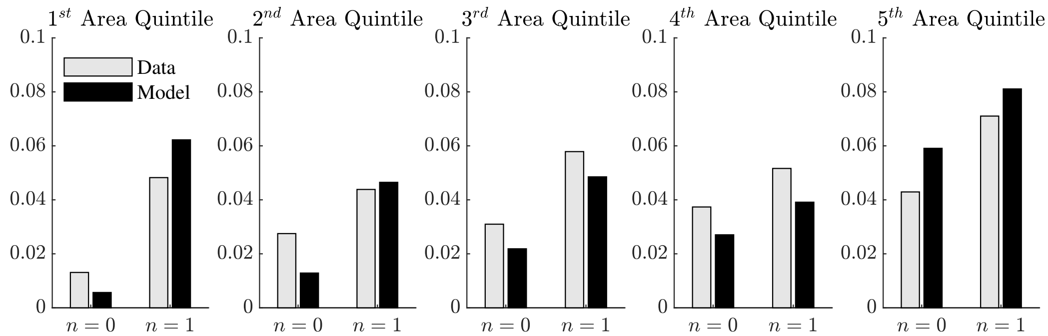
As a validation exercise, we simulate data from the estimated model to compute the partial correlation between d_{it} and $\mathcal{N}_{it}$ and compare it to the estimated coefficient β_1 in the drilling regression (cf., equation 8 in Section 3.3). Recall that this "reduced-form" parameter captures the strength of strategic substitutability between neighbors' drilling decisions. Starting at a steady state on each of 10 replications of the cadastral village maps, we simulate five-year panels consisting of triplets $\{d_{it}, n_{it}, \mathcal{N}_{it} : t = 1, ..., 5\}$ for every plot on the map that is assigned developed status (see Step 2 in Subsection 5.3); this yields 43,702 5-year panels in total. Using this very large synthetic data "sample", we estimate a linear probability model version of the drilling reduced form (corresponding to column 5 of Appendix Table F.1), essentially giving a "population" value of β_1 from the model.[41] Figure 6 shows

[39] For instance, using differences in log odds ratios, we obtain $\theta/\sigma = \log \frac{P_{0,k}(1-P_{0,k'})}{P_{0,k'}(1-P_{0,k})}[f(1, a_k; \alpha, \delta) - f(1, a_{k'}; \alpha, \delta)]^{-1}$.

[40] These standard errors are understated because we do not account for pre-estimated first stage parameters. Given the high precision, however, any correction for first stage sampling error is unlikely to matter for inference.

[41] Because these synthetic reference plots are each endowed with unobserved flow/fail heterogeneity (see Step 1 in Subsection 5.3), we use reference plot fixed effects just as with the real data. However, since misclassification error is not

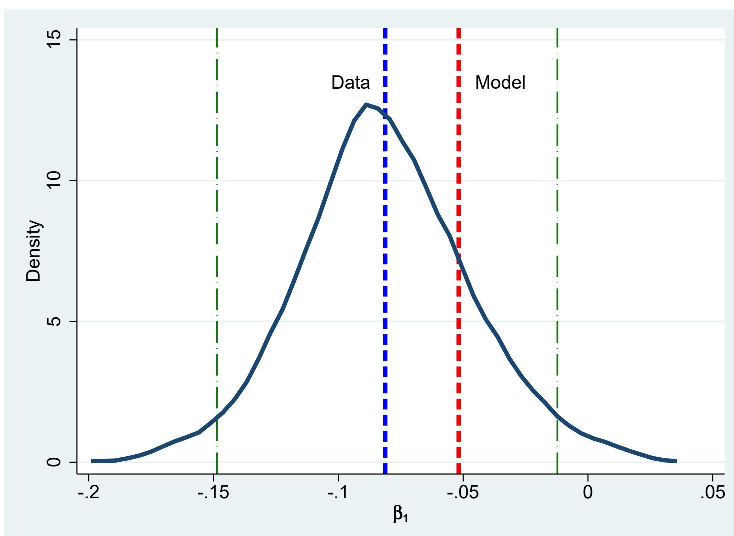
Figure 5: Annual drilling rates by plot area and n



Notes: Each pair of bars represents a data moment (annual drilling rate by plot area quintile and number of functioning wells on the reference plot) and its corresponding model fit.

the bootstrap distribution of $\hat{\beta}_1$ based on the actual data with the 95% confidence interval marked by the vertical green lines. Corroborating the model, we find that the "population" value of β_1 implied by the estimated structural parameters (vertical red line) lies well within the confidence interval.

Figure 6: Strategic interactions – data vs. model



Notes: Estimate of strategic interaction parameter β_1 (vertical blue line) and its bootstrapped distribution from a linear fixed-effects regression of a drilling indicator on the number of functioning wells outside the reference plot (column 5 of Appendix Table F.1). Vertical green lines denote 95% confidence interval bounds for $\hat{\beta}_1$. Vertical red line denotes the "population" value of β_1 implied by the structural model.

an issue in the simulated data, we do not use IV. Finally, since our model assumes no drilling on plots with two borewells, we drop observations with $n = 2$ in the estimation samples for both actual and synthetic data.

6 Counterfactuals

Our quantitative policy evaluation addresses three inter-related questions: (1) What is the social cost of the current policy of free (but rationed) electricity to farmers for pumping groundwater? (2) What is the optimal tax on borewells that eliminates both the deadweight losses due to electricity subsidies and the well interference externality? (3) How should policy treat the existing stock of borewells?

6.1 Social value of groundwater development

Based on our estimates, the (marginal) private value of developed land is 77,000 Rs/acre.[42] The social value of groundwater development is this private value minus the cost of electricity, which, though given free to borewell owners, is not free to society. We find that this social value is only about 31,000 Rs/acre. In other words, 60 percent of the private value of groundwater development is accounted for by the capitalized value of the electricity subsidy. Next, we recompute the cadastral map equilibria in a counterfactual economy wherein each and every plot is an island unto itself, thereby zeroing out interference effects between borewells operating in adjacent plots. Compared to the current equilibrium, we find that the borewell density in steady state increases by almost 30 percent (from 0.28 to 0.36 wells/acre) and that the social value of groundwater development rises to 55,000 Rs/acre. Thus, the negative externality diminishes the value of groundwater to society by a substantial 44 percent.

6.2 Policy analysis

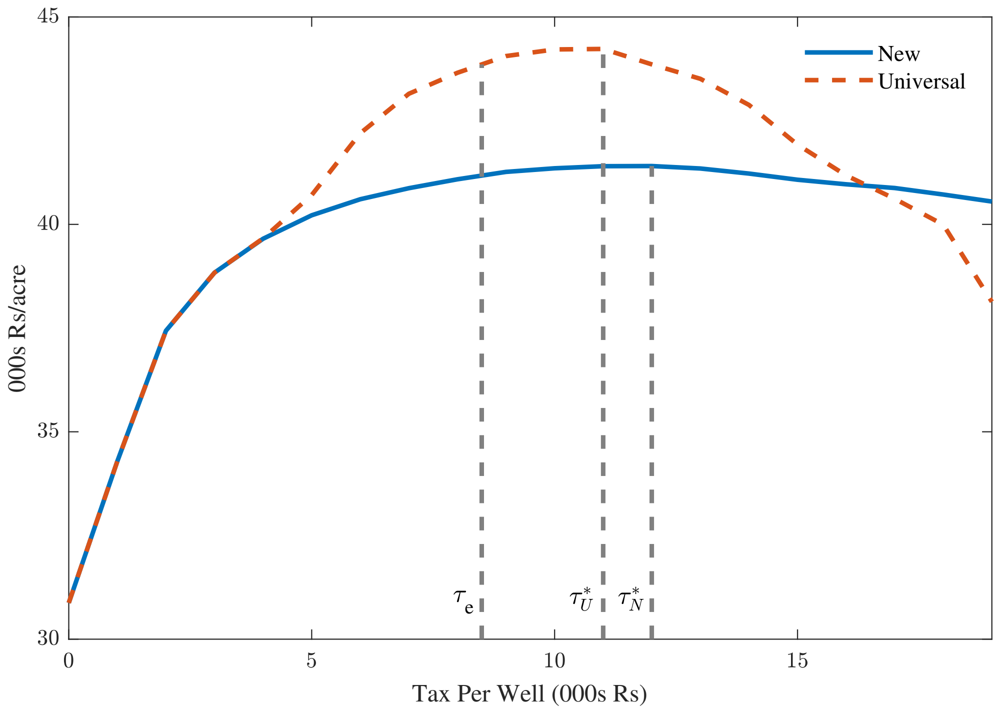
We now assess the social welfare implications of an annual tax τ on functioning borewells. In practice, τ could be implemented as a flat charge for maintaining an agricultural electrical connection. Setting τ for all functioning borewells equal to the annual cost of electricity to run the pump given daily power rationing (i.e., τ_e =8,500 Rs/year) would fully recover costs from agricultural consumers. A flat charge exceeding τ_e would act like a Pigouvian tax on borewells.[43]

For each hypothetical value of the borewell tax τ, we compute social welfare along the transition path from the zero tax baseline (see Appendix H for details on the equilibrium concept and solution algorithm). This calculation takes into account the "short-run", over which the existing stock of borewells is relevant and, therefore, allows us to compare different policy treatments of wells that have already

[42] We define this private value as the average expected discounted present value of agricultural output minus drilling costs in steady state across all maps weighted by the proportion of total acreage from the map village represented in the sample. To be clear, the private value accounts for externalities inasmuch as it averages across hundreds of adjacent plots with (potentially) mutually interfering borewells.

[43] In terms of the model, the annual *net* Rupee value of output under a counterfactual $\tau > 0$ becomes $\mathbb{E}\, y - \tau n$. Once a borewell fails, its owner incurs no further tax on it.

Figure 7: Social welfare under alternative borewell taxes



Notes: Each point on the solid (dashed) curve represents the social welfare along the transition path from the benchmark zero-tax economy to the long-run steady state under a hypothetical tax on newly drilled (both new and old) borewells: $\tau_e = 8.5$ is the tax that recovers electricity costs; $\tau_D^* = 11.0$ is the optimal universal tax; $\tau_N^* = 12.0$ is the optimal tax on newly drilled borewells.

been sunk. In particular, we consider two alternative tax regimes: The first is a universal tax that applies to all borewells, both those sunk before and after the policy is implemented. We assume that extant borewells, once subject to taxation, can be dismantled at zero cost.[44] Under the second tax regime, extant wells are grandfathered so that the tax only applies to newly drilled borewells and there is no dismantling.

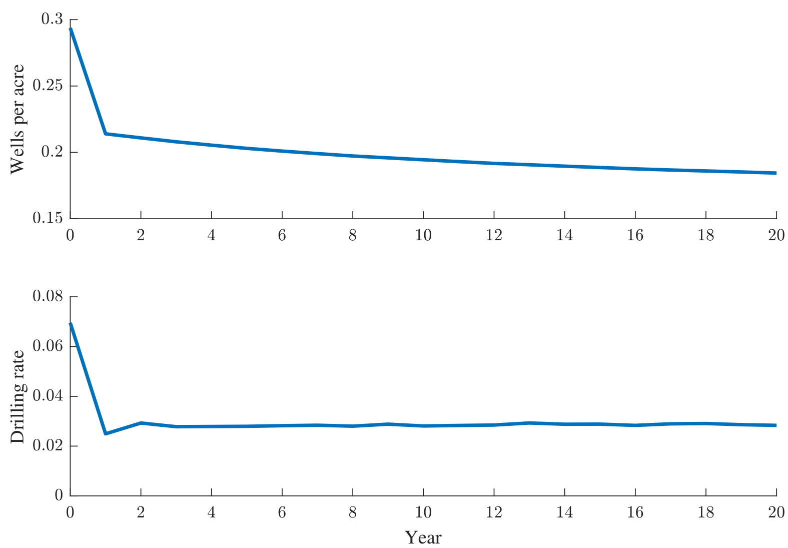
Electricity cost recovery: We first consider the implications of levying a tax $\tau = \tau_e$, equal to the annual cost of electricity, on all functioning borewells. Given the prohibitive cost of metering hundreds of thousands of borewells, Shah et al. (2007) propose similarly structured electricity charges combined with quantity rationing (see also Shah et al. 2012). As shown in Figure 7, removing the electricity

[44]Although our structural model does not explicitly incorporate dismantling (which would never be chosen anyway), our counterfactuals still treat this decision as the outcome of a strategic equilibrium (Appendix H). In particular, we assume that once the universal tax is implemented, borewell owners adopt beliefs (one-period ahead state transition probabilities) consistent with the transition to the new steady state under the counterfactual policy. Thus, in making their dismantling decision, each borewell owner takes into account the dismantling in their own adjacency.

subsidy through such a mechanism would increase the social value of groundwater development from 31,000 to 43,000 Rs/acre, or by 39 percent. In other words, the deadweight loss from free electricity provision, and the consequent over-drilling, is 12,000 Rs (170 US$) per developed acre in present value terms, or about 3% of (developed) farmland value. By comparison, the steady-state fiscal cost of the subsidy in present value terms is about 48,000 Rs/acre (= 0.28 borewells per acre $\times$ 8500/0.05, where the second term is τ_e divided by the discount rate).

Figure 8 illustrates the behavioral responses underlying these social gains. As soon as the policy comes into force (year 1), we see a substantial (20%) dismantling of extant borewells accompanied by a big (70%) drop in the annual drilling rate. The rather slow decline in borewell density over the next 20 years of transition to the new steady-state, from 0.22 to 0.18 wells per acre, is due to selective dismantling; pre-existing borewells that continue to function after the tax is imposed are largely of the high flow/low failure type and hence are relatively durable.

Figure 8: Transition dynamics under electricity cost recovery



Notes: The top (bottom) graph shows the transition dynamics for borewell density (annual drilling rate) to the new steady state under an annual borewell tax implemented in year 1 that recovers electricity costs $\tau = \tau_e$.

Optimal borewell taxation: The social welfare maximizing tax τ^* should exceed the annual cost of electricity to correct the negative externality. We see in Figure 7 that the optimal universal tax, τ_U^*, is 11,000 Rs, which represents a 23% "Pigouvian premium" over electricity costs. When only newly drilled wells are subject to taxation, however, the optimal tax, τ_N^*, is 12,000 Rs, a 35% Pigouvian premium; the higher premium is required because the marginal externality cost of a new borewell is greater when there is no dismantling of extant borewells. Nevertheless, the extra 1,000 Rs tax implied by the policy of grandfathering extant borewells yields only a small social gain.[45] Moreover, if we compare social welfare under the two tax regimes at τ_U^*, the difference is only about 3,000 Rs/acre. This surprisingly small amount reflects the limited social gains achieved from dismantling nearly one-third of pre-existing borewells, which, while no longer viable under the universal tax, generated a revenue flow more than justifying their drilling costs ex-ante. We find, therefore, that a tax of 11,000 Rs per borewell per year *applied only to borewells sunk ex-post* (and implemented, e.g., by charging an annual flat-rate only for new electrical connections) achieves nearly maximal social welfare along the transition path to the new steady state. This is a felicitous result because, by restricting the tax to new borewells, no farmer would suffer a capital loss on sunk investments, making this policy far more politically palatable than a universal borewell tax at only a small cost in terms of forgone social wefare.

7 Conclusion

We set out to assess the social cost of a policy of free electricity to farmers for groundwater pumping in South India, a context with economically important yet highly localized externalities. In so doing, we developed a tractable dynamic strategic equilibrium model of borewell investment across a large network of agricultural plots along with a novel simulation-based estimation strategy. Our general approach is potentially applicable to a wide range of settings where spatial spillovers are salient.

We find that the social cost of free electricity amounts to 39% of the value of groundwater development, a deadweight loss of US$170 per acre of land with groundwater potential, or around 3% of its market value. Given the externality, we also address the broader question of the optimal tax on borewells, one which would not only eliminate the extensive margin distortion from the electricity subsidy but which would also maximize social welfare. In accounting for how such a tax affects existing borewell investments in the transition to the new steady state, our counterfactual analysis is uniquely able to capture the "short-run", relevant to borewell owners and policymakers alike. Indeed, adopting a social welfare criterion (and thereby setting aside cost-recovery considerations), we find that the most sensible policy is to tax only newly drilled borewells, charging a modest Pigouvian premium above and beyond the annual cost of electricity. With minimal loss in social welfare relative to a universal tax on borewells, this policy avoids a capital levy on existing well owners, an influential political interest group in rural India.

[45] Intuitively, the steady-state annual drilling rate at τ =11,000 is already quite low at 0.02 (down from around 0.07 at baseline). An increase in τ, therefore, will not induce much behavioral response and will thus have little impact on social welfare at the margin.

References

ACEMOGLU, D., C. GARCIA-JIMENO, AND J. A. ROBINSON (2015): "State capacity and economic development: A network approach," *American Economic Review*, 105, 2364–2409.

AGUIRREGABIRIA, V. AND P. MIRA (2010): "Dynamic discrete choice structural models: A survey." *Journal of Econometrics*, 156, 38–67.

ARMANTIER, O. AND N. TREICH (2009): "Subjective probabilities in games: An application to the overbidding puzzle," *International Economic Review*, 50, 1079–1102.

BADIANI-MAGNUSSON, R. AND K. JESSOE (2018): "Electricity Prices, Groundwater, and Agriculture: The Environmental and Agricultural Impacts of Electricity Subsidies in India," in *Agricultural Productivity and Producer Behavior*, University of Chicago Press, 157–183.

BAJARI, P., V. CHERNOZHUKOV, H. HONG, AND D. NEKIPELOV (2015): "Identification and efficient semiparametric estimation of a dynamic discrete game," Tech. rep., National Bureau of Economic Research.

BENKARD, C. L., P. JEZIORSKI, AND G. Y. WEINTRAUB (2015): "Oblivious equilibrium for concentrated industries," *The RAND Journal of Economics*, 46, 671–708.

BLAKESLEE, D., R. FISHMAN, AND V. SRINIVASAN (2020): "Way down in the hole: Adaptation to long-term water loss in rural India," *American Economic Review*, 110, 200–224.

CHANDRAKANTH, M. (2015): *Externality in Irrigation*, New Delhi: Springer India, 25–36.

COLE, S., X. GINÉ, J. TOBACMAN, P. TOPALOVA, R. TOWNSEND, AND J. VICKERY (2013): "Barriers to household risk management: Evidence from India," *American Economic Journal: Applied Economics*, 5, 104–35.

DOMEIJ, D. AND J. HEATHCOTE (2004): "On the distributional effects of reducing capital taxes," *International economic review*, 45, 523–554.

ECKSTEIN, Z. AND K. I. WOLPIN (1999): "Why youths drop out of high school: The impact of preferences, opportunities, and abilities," *Econometrica*, 67, 1295–1339.

ERICSON, R. AND A. PAKES (1995): "Markov-perfect industry dynamics: A framework for empirical work," *The Review of economic studies*, 62, 53–82.

Fafchamps, M. and J. Pender (1997): “Precautionary saving, credit constraints, and irreversible investment: Theory and evidence from semiarid India.” *Journal of Business & Economic Statistics*, 15, 180–194.

Fishman, R., X. Giné, and H. G. Jacoby (2023): “Efficient irrigation and water conservation: Evidence from South India,” *Journal of Development Economics*, 103051.

Fishman, R. M., T. Siegfried, P. Raj, V. Modi, and U. Lall (2011): “Over-extraction from shallow bedrock versus deep alluvial aquifers: Reliability versus sustainability considerations for India’s groundwater irrigation,” *Water Resources Research*, 47.

Fu, C. and J. Gregory (2019): “Estimation of an Equilibrium Model With Externalities: Post-Disaster Neighborhood Rebuilding,” *Econometrica*, 87, 387–421.

Giné, X. and H. G. Jacoby (2020): “Contracting under uncertainty: Groundwater in South India,” *Quantitative Economics*, 11, 399–435.

Gulati, M. and S. Pahuja (2012): *Direct delivery of power subsidy to agriculture in India*, Sustainable Energy for All.

Hora, T., V. Srinivasan, and N. B. Basu (2019): “The groundwater recovery paradox in South India,” *Geophysical Research Letters*, 46, 9602–9611.

Huang, L. and M. D. Smith (2014): “The dynamic efficiency costs of common-pool resource exploitation.” *American Economic Review*, 104, 4071–4103.

Ifrach, B. and G. Weintraub (2017): “A framework for dynamic oligopoly in concentrated industries,” *Review of Economic Studies*, 1106–1150.

Jacoby, H. G. (2017): ““Well-fare” Economics of Groundwater in South Asia.” *World Bank Research Observer*, 32, 1–20.

Jain, M., R. Fishman, P. Mondal, G. L. Galford, N. Bhattarai, S. Naeem, U. Lall, Balwinder-Singh, and R. S. DeFries (2021): “Groundwater depletion will reduce cropping intensity in India,” *Science advances*, 7, eabd2849.

Katic, P. G. and R. Q. Grafton (2012): “Economic and spatial modelling of groundwater extraction,” *Hydrogeology Journal*, 20, 831–834.

König, M. D., D. Rohner, M. Thoenig, and F. Zilibotti (2017): "Networks in conflict: Theory and evidence from the great war of africa," *Econometrica*, 85, 1093–1132.

Kumar, M. D., M. Sivamohan, V. Niranjan, and N. Bassi (2011): "Groundwater management in Andhra Pradesh: time to address real issues," *Occasional paper*, 4.

Lin, C. Y. C. (2013): "Strategic Decision-Making with Information and Extraction Externalities: A Structural Model of the Multi-Stage Investment Timing Game in Offshore Petroleum Production." *Review of Economics and Statistics*, 95, 1601–1621.

Magnac, T. and D. Thesmar (2002): "Identifying dynamic discrete decision processes," *Econometrica*, 70, 801–816.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 60, 531–542.

Pfeiffer, L. and C. Y. C. Lin (2012): "Groundwater pumping and spatial externalities in agriculture." *Journal of Environmental Economics and Management*, 64, 16–30.

Rust, J. (1994): "Structural estimation of Markov decision processes," *Handbook of econometrics*, 4, 3081–3143.

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

Sayre, S. S. and V. Taraz (2019): "Groundwater depletion in India: Social losses from costly well deepening," *Journal of Environmental Economics and Management*, 93, 85–100.

Sears, L., C.-Y. C. L. Lawell, and M. T. Walter (2022): "Groundwater Under Open Access: A Structural Model of the Dynamic Common Pool Extraction Game," *unpublished manuscript, Cornell University.*

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

Shah, T. (2010): *Taming the anarchy: Groundwater governance in South Asia*, Routledge.

Shah, T., M. Giordano, and A. Mukherji (2012): "Political economy of the energy-groundwater nexus in India: exploring issues and assessing policy options," *Hydrogeology Journal*, 20, 995–1006.

Shah, T., C. Scott, and A. K. A. Sharma (2007): "11 Energy–Irrigation Nexus in South Asia: Improving Groundwater Conservation and Power Sector Viability," *THE AGRICULTURAL GROUNDWATER REVOLUTION*, 211.

Sidhu, B. S., M. Kandlikar, and N. Ramankutty (2020): "Power tariffs for groundwater irrigation in India: A comparative analysis of the environmental, equity, and economic tradeoffs," *World Development*, 128, 104836.

Weintraub, G. Y., C. L. Benkard, and B. Van Roy (2008): "Markov perfect industry dynamics with many firms," *Econometrica*, 76, 1375–1411.

Xu, H. (2018): "Social interactions in large networks: A game theoretic approach," *International Economic Review*, 59, 257–284.

Appendix (Not for Publication)

Contents

A Additional figures and tables **41**

B Exploratory regressions: Developed status and drilling **44**

C Average annual electricity costs per borewell **45**

D Cadastral map villages **45**

E Well failure and duration dependence **48**

F Drilling propensity: Linear FE-IV estimates **49**

G Land values and developed status **51**

H Transitional dynamics **53**

A Additional figures and tables

Figure A.1: Monsoon rainfall at mandal level by year

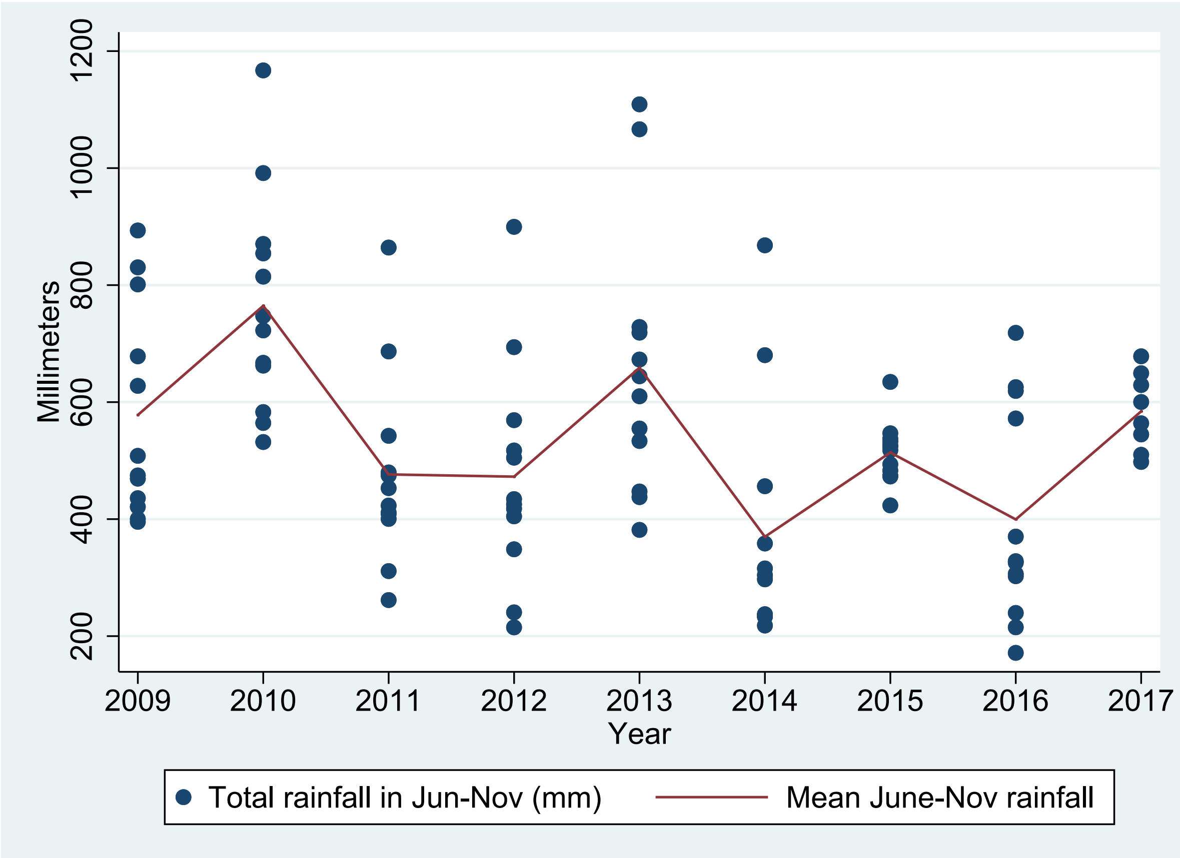
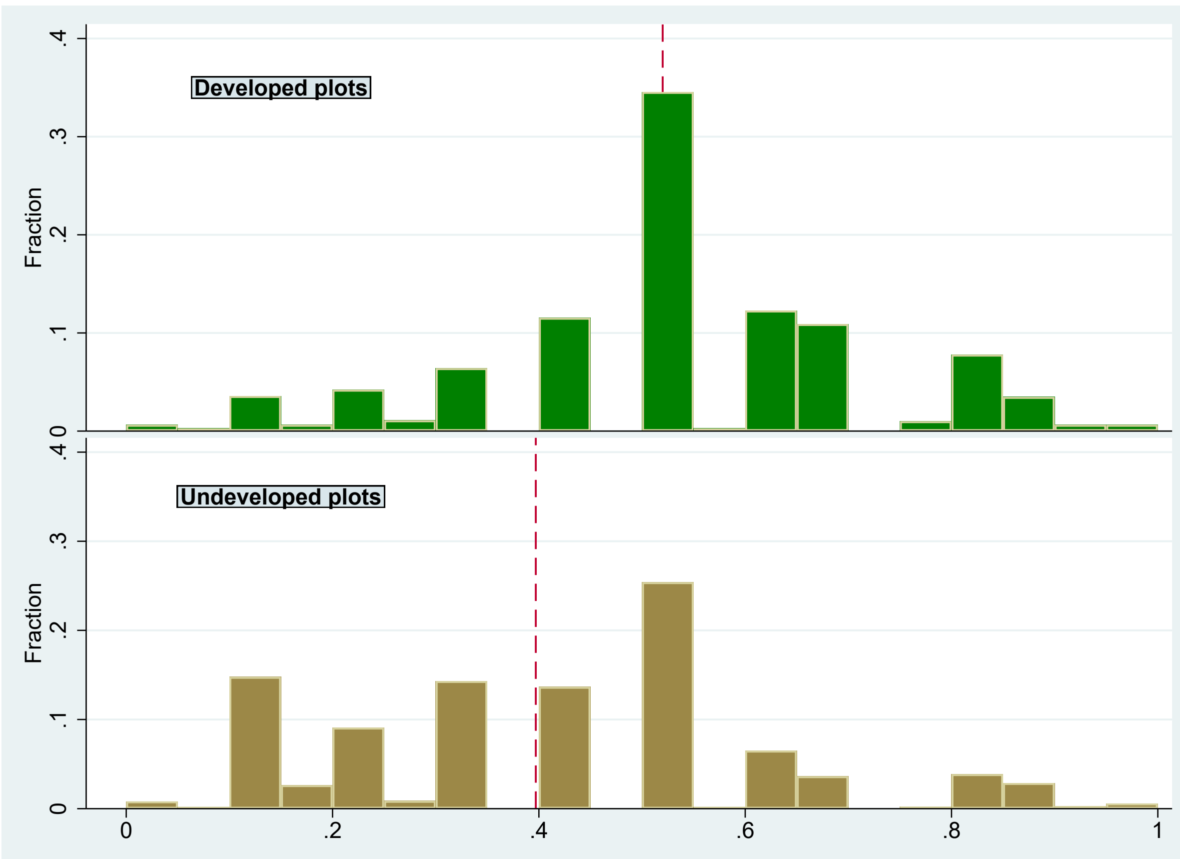


Figure A.2: Drilling success on developed and undeveloped plots



Notes: Vertical dashed lines indicate means of subjective probability of success for, respectively, developed and undeveloped plots.

Table A.1: Well Failure by Year

Year	Functional	Failed	Total
2012	559	47	606
	(92.2)	(7.8)	(100)
2013	556	41	597
	(93.1)	(6.9)	(100)
2014	527	53	580
	(90.9)	(9.1)	(100)
2015	512	33	545
	(93.9)	(6.1)	(100)
2016	489	34	523
	(93.5)	(6.5)	(100)
Total	2,643	208	2,851
	(92.7)	(7.3)	(100)

Notes: Percent of yearly total in parentheses. Sample consists of reference plot borewells subject to failure in each year.

Table A.2: End-of-Season Well Flow

	Frequency (%)	
Flow	2010	2017
0.10	32	114
	(6.2)	(22.2)
0.25	57	219
	(11.1)	(42.6)
0.50	172	143
	(33.5)	(27.8)
0.75	192	35
	(37.4)	(6.8)
1.00	61	3
	(11.9)	(0.6)
Total	514	514
	(100)	(100)
Mean	0.600	0.325
Std. dev.	0.245	0.193

B Exploratory regressions: Developed status and drilling

We explore the determinants of plot development status ($\mathcal{D}$) as well as drilling conditional on the plot being developed. Each logit regression in Table B.1 includes both plot area quintile and village dummies (for the 44 sample villages). We also include, from the 2010 household survey, a measure of each plot owner's gross wealth defined as the total value of household assets as of 2009, including agricultural land, livestock, agricultural machinery, household durable goods, and savings in the form of bank deposits, cash and jewelry. Total pre-sample wealth should be a good proxy for liquidity.

In column 1, we see that wealthier households (in 2009) are significantly more likely to have developed plots (i.e., with functioning borewells and/or drilling activity during 2012-2016). This apparent wealth or liquidity effect is accentuated once we control for the subjective probability of drilling success in 2010 (col. 2), which itself strongly predicts developed status, and with the expected sign. By contrast, we find no significant association between pre-sample wealth and the propensity to drill on developed plots, whether or not we control for the probability of drilling success (compare cols. 3 and 4).[1]

Discussion It is important to stress that these liquidity effects (or lack thereof) are associations, i.e., not necessarily causal. For example, farmers with developed plots may have had higher agricultural surplus in the past and have thus been able to accumulate more wealth prior to 2012-16. Similarly, suitability for groundwater development not captured by the subjective probability of drilling success may, through past groundwater use, partly determine past wealth accumulation. Such correlations, however, cannot explain the zero liquidity effects found in columns 3 and 4 of Table B.1.

Further, according to our 2017 survey of borewell owners, only 27% of respondents relied on their own savings as the main source of finance for the largest components of the cost of borewell investment, drilling the bore and purchase of the submersible pump. Most farmers use various forms of formal and informal credit. More broadly, data from the household credit module contained in the 2010 household survey show that, out of the 1488 respondents, 89% had outstanding bank credit, 46% had loans from family or friends, and 45% were borrowing from moneylenders. These percentages are very similar across households that have at least one borewell on their land and those that have none. This substantial credit access in our setting may explain why pre-sample wealth is uncorrelated with subsequent drilling and provides empirical justification for our borewell investment model (on developed plots) that abstracts from financial frictions.

[1] In these drilling regressions, we also control for the initial (i.e., 2011) number of functioning borewells.

Table B.1: Determinants of developed status and drilling

	$\Pr(\mathcal{D})$		$\Pr(drill \mid \mathcal{D}=1)$	
	(1)	(2)	(3)	(4)
Log(gross wealth in 2009)	0.139***	0.181***	0.0878	0.0778
	(0.0510)	(0.0527)	(0.0593)	(0.0594)
Prob(drilling success 2010)	—	2.904***	—	-0.461*
		(0.223)		(0.257)
No. functioning borewells 2011	—	—	-1.034***	-1.007***
			(0.0835)	(0.0847)
Fixed effects	village	village	village	village
Log-likelihood	-1726.9	-1633.4	-1600.9	-1599.3
Observations	2,862	2,862	5,540	5,540
Number of villages	44	44	44	44

Notes: Standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$). All regressions by (village) random effects logit. Dependent variable in columns 1-2 is developed status of plot (see main text, Section 2.2) and in columns 3-4 is whether a drilling attempt was made on plot during the year (from 2012-16). Each regression also controls for plot area quintile dummies (coefficients not reported).

C Average annual electricity costs per borewell

The electricity cost of a borewell per year is the product of (1) power consumption of the average pump of 6 horsepower (HP), which is 4.5 kWh (= 6 HP×0.746 kWh/HP), (2) 630 annual hours of pumping (average of unit record data for our 12 sample mandals from India's 4th Minor Irrigation Census), and (3) marginal cost of electricity of 3 Rs/kWh (Gulati and Pahuja, 2012). All three components in this calculation are likely overly conservative estimates, so that 8500 Rs. should be viewed as a lower bound on the true electricity cost.

D Cadastral map villages

The villages for which we have cadastral maps are Pamireddypalli in Atmakur mandal, Dharmapur and Ramachandrapuram in Mahabubnagar mandal, Jajapur in Narayanapet and Thipparasipalli in Utkur madal. In Anatapur district, we have cadastral maps for Manesamudram in Hindupur mandal, M. Venkata Puram and Manepalli, both part of the same panchayat in Lepakshi mandal, Y.B. Halli in Madakasira Muddireddy Palli in Parigi Chalakuru and Somandepalli, both part of the same panchayat in Somandepalli mandal, Siddarampuram and Reddipalli in B.K. Samudram mandal, Itukalapalli in

Anantapur and Ayyavaripalli in Rapthadu mandal.

We take these maps to be representative of all villages for which we have data in each respective mandal. We use the digitized maps, as in the example shown in Figure D.1, to create 14 plot adjacency matrices defining the networks upon which the dynamic discrete investment game is played.

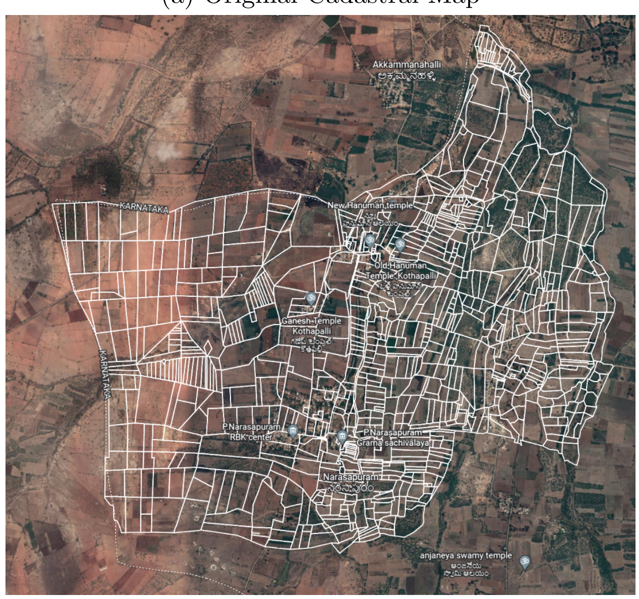
Table D.1: Predicted fraction of developed plots by map village

	Plot area quintiles					
Map village	1st	2nd	3rd	4th	5th	All
Ayyavaripalli	0.108	0.175	0.226	0.227	0.298	0.222
Dharmapur	0.341	0.475	0.554	0.557	0.644	0.533
Itukalapalli	0.195	0.298	0.368	0.370	0.459	0.321
Jajapur	0.259	0.380	0.457	0.459	0.550	0.433
M. Venkata Puram	0.155	0.243	0.306	0.308	0.391	0.254
Manesamudram	0.187	0.287	0.356	0.358	0.446	0.276
Muddireddy Palli	0.198	0.302	0.373	0.375	0.464	0.276
Pamireddypalli	0.365	0.501	0.580	0.582	0.668	0.605
Ramachandrapuram	0.435	0.574	0.650	0.652	0.730	0.603
Reddipalli	0.238	0.354	0.429	0.431	0.522	0.396
Siddarampuram	0.442	0.581	0.656	0.658	0.735	0.635
Suddakuntapalli	0.135	0.215	0.273	0.275	0.354	0.220
Thipparasipalli	0.273	0.397	0.475	0.477	0.568	0.475
Y.B. Halli	0.154	0.242	0.305	0.307	0.390	0.263
All map villages	0.210	0.350	0.435	0.433	0.547	0.387

Notes: Predicted probability of developed plot from logit regression of $\mathcal{D}$ on plot area quintile and map village dummies.

Figure D.1: Village Muddiredipalle

(a) Original Cadastral Map



(b) Digitized Cadastral Map

E Well failure and duration dependence

A simple test of duration dependence in well failure that avoids the intricate specification issues of duration modelling is to check whether the probability of failure between 2012-16 is related to well age in 2012, which is predetermined. The results in Table E.1 indicate significant duration dependence. The marginal effect from the column 1 estimates implies that a well that was 10 years older in 2012 has a failure rate 0.092 higher over the subsequent five years. All of this effect, however, appears to be concentrated among the 59 wells that were more than 20 years old in 2012 (see, especially, column 3).

Table E.1: Well Age and Subsequent Failure

	(1)	(2)	(3)
Age in 2012	0.0428***	—	—
	(0.0130)		
Age$\times\mathbb{1}_{Age\leq 10}$	—	0.00545	—
		(0.0319)	
(Age-10)$\times\mathbb{1}_{10<Age\leq 20}$	—	0.0375	—
		(0.0372)	
Age$\times\mathbb{1}_{Age\leq 20}$	—	—	0.0165
			(0.0169)
(Age-20)$\times\mathbb{1}_{20<Age}$	—	0.125***	0.132***
		(0.0456)	(0.0466)
Observations	606	606	606
log-likelihood	-375	-375.3	-375.5
Equal slopes test (p-value)	—	0.028	0.006

Notes: Standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$). Dependent variable is indicator for whether well failed between 2012-16. Estimation is by ML logit. Constant term not reported. Test of equal slopes compares spline coefficients (3 in column 2 and 2 in column 3).

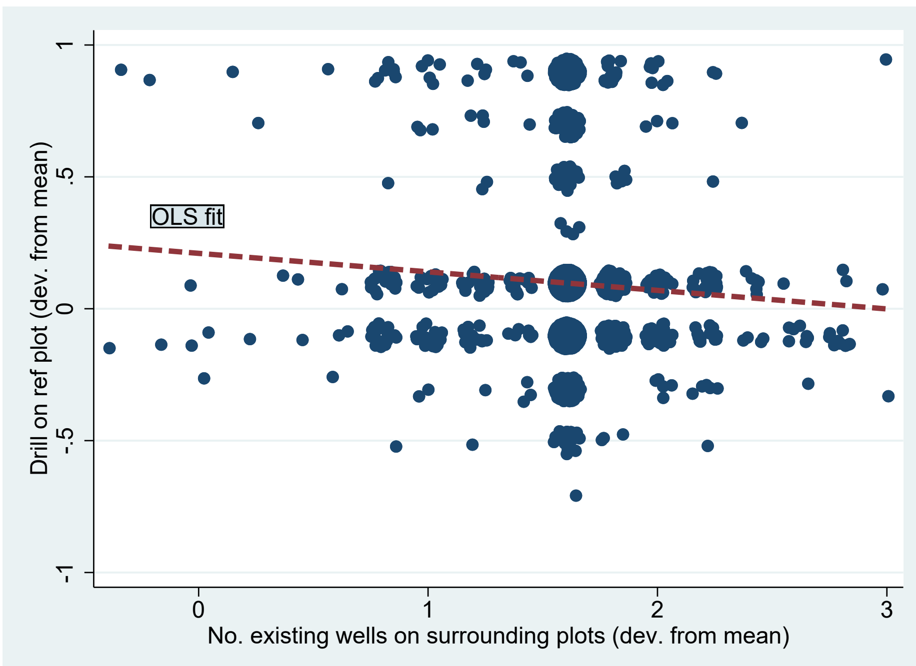
F Drilling propensity: Linear FE-IV estimates

Using the notation from the main text, we estimate a linear probability model for drilling of the form

$$d_{it} = \alpha_i + \beta_1 \mathcal{N}_{it} + \beta_2 R_{mt-1} + \varepsilon_{it}, \tag{F.1}$$

now treating α_i as a fixed (rather than random) effect. We assume measurement error of the classical variety and use the number of existing wells in the adjacency (outside the reference plot), $\mathcal{N}_{it}^E$, as an instrument. Figure F.1 shows the within reference plot (i.e., fixed effects) regression of d_{it} on $\mathcal{N}_{it}^E$, which is essentially the reduced form corresponding to our IV regression of equation F.1.

Figure F.1: Drilling and the number of existing wells in the adjacency



Column (1) of Table F.1 reports the least-squares (FE) coefficient ignoring measurement error. As in Table 1 of the main text, we do not find a significant impact of neighboring wells on drilling. Column (2) shows the first stage regression of $\mathcal{N}_{it}$ on the instrument $\mathcal{N}_{it}^E$ and column (3) the resulting FE-IV estimate. Just as with the CRE-MEM estimator in Table 1, here we find a significantly negative effect of neighboring wells once we correct for measurement error. One concern, however, is that, if there is spatial correlation in the unobservables, then $\mathcal{N}_{it}^E$ may be correlated with the residuals, which contain the effect of *own* borewells on drilling. To assess this, in column (4) we add dummies for the number of borewells on the reference plot to remove the effect of own borewells from the residuals.[2] That there is no appreciable difference between the estimates of β_1 across columns (3) and (4) gives us further confidence that negative effect of neighboring wells on reference plot drilling is indeed causal.

Table F.1: Determinants of drilling 2012-16–Linear probability models

	(1) drill	(2) $\mathcal{N}$	(3) drill	(4) drill	(5) drill
No. func wells exc ref plot ($\mathcal{N}$)	-0.0215		-0.0815**	-0.0820**	-0.0812**
	(0.0154)		(0.0343)	(0.0321)	(0.0344)
Good monsoon (R)	-0.00528	0.00798	-0.00367	-0.00379	
	(0.0135)	(0.01000)	(0.0136)	(0.0133)	
No. exist wells exc ref plot		0.859***			
		(0.0416)			
1 func well on ref plot				-0.238***	
				(0.0246)	
2 func wells on ref plot				-0.441***	
				(0.0543)	
Reference plot FE	YES	YES	YES	YES	YES
Observations	3,815	3,815	3,815	3,815	3,367
R^2	0.203	0.968	-0.004	0.062	-0.006

Notes: Standard errors in parentheses clustered by reference plot (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$). Columns 1 and 2 are by least-squares with fixed effects; columns 3-5 are by two stage least squares using the number of existing wells in adjacency (outside of reference plot) as instrument. Column 5 drops observations with more than one functioning well on the reference plot.

[2] Insofar as past drilling successes lead to more borewells on the reference plot, the fixed effects estimator of the own borewell coefficients in this short panel are biased (akin to Nickell bias). We thus refrain from comparing the relative magnitudes of own borewell and neighboring borewell coefficients between Table F.1 and Table 1 in the main text.

G Land values and developed status

We use plot value data collected in the 2017 household survey to estimate the difference in present discounted values between developed and undeveloped land. Specifically, the survey asked each plot owner "if you were to sell this plot today, including the associated water rights, how much would you receive in thousands of Rs. per acre?" In evaluating the present discounted value of the projected income flows off of their land, as reflected in their stated sales price, we presume that farmers use the same discount factor that they would use in assessing the future net benefits from drilling. This presumption is the basis for our calibration of β for the structural estimation.

To estimate the average marginal value of a developed plot, we regress the reported value per acre of plot j on the developed status indicator $\mathcal{D}_j$. Needless to say, $\mathcal{D}_j$ is potentially endogenous. For instance, unobserved land amenities (e.g., ready access to markets, good soil) may both increase land values and encourage groundwater development. It is also plausible that liquidity constrained households both own less valuable land and can less afford to develop their land for groundwater extraction. To deal with such reverse causality, we focus on households owning multiple plots in the same village and use household fixed effects to estimate the land value regression. This procedure controls for both unobservable household-specific and location specific factors.

In Table G.1, we report three regressions on each of two estimation samples. The first ("full") sample (cols. 1-3) consists of all plots with non-missing land values that are owned by multi-plot households. The second ("trimmed") sample (cols. 4-6) drops land value observations below the 5th and above the 95th percentiles and is restricted to households owning at least two plots with admissible land values. For each sample, we report, respectively, an OLS, village fixed effects, and household fixed effects regression.

The village fixed effects estimator, based on 44 sample villages, purges locational factors correlated with both land values and developed status at the village level. That the coefficient on developed status does not fall (it actually rises a bit) in moving from OLS to village fixed effects indicates that these unobserved location characteristics are not a serious confound. Similarly, the finding that the coefficient on developed status changes little in moving from village to household fixed effects (especially in the trimmed sample) suggests that liquidity constraints, insofar as they determine development status (see Appendix B), are not strongly correlated with plot values. Unsurprisingly, standard errors are much smaller with the trimmed sample than with the full sample. Even so, the household fixed effects estimates of the average marginal value of a developed plot are virtually identical across samples at around 80,000 Rs/acre, representing a 25% market premium over an undeveloped plot.

Finally, we note a threat to the validity of our household fixed effect estimator: unobserved *plot-level* characteristics (e.g., soil quality) correlated with both land values and developed status. Recall,

Table G.1: Plot values and developed status

	Full sample			Trimmed sample		
	(1)	(2)	(3)	(4)	(5)	(6)
Developed ($\mathcal{D}=1$)	94.78	98.98	81.69	66.46	82.45	79.30
	(19.88)	(17.08)	(16.55)	(12.59)	(9.402)	(6.516)
Observations	2,346	2,346	2,346	2,093	2,093	2,093
R^2	0.011	0.014	0.021	0.033	0.064	0.130
Fixed effects	none	village	household	none	village	household
No. of clusters	44	44	898	44	44	804

Notes: Robust-clustered standard errors in parentheses. Dependent variable is plot value in thousands of Rs. per acre. Full sample (cols. 1-3) consists of all plots owned by multi-plot households. Trimmed sample (cols. 4-6) removes land value observations below the 5th and above the 95th percentiles and is restricted to households owning at least two plots with admissible land values. Constant term not reported.

however, that location-specific unobservables, a far more important component of residual variation across villages than across household plots *within* villages, have little impact on our regression results. This finding suggests that any bias due to unobserved plot-level characteristics is likely to be negligible.

H Transitional dynamics

H.1 Equilibrium

We now describe the adjacency equilibrium over the transition path of the benchmark village economy to a new steady state following a policy reform at date $t = 1$ which introduces a tax τ on borewells. The village network transits to a new (steady-state) adjacency equilibrium AE_τ. We assume that i) CCPs along the transition depend only on the state of the adjacency and date t, and ii) that the plot owner has beliefs about the evolution of $X_{(i1)t}$ along an "average" transition. Assumption ii) requires that equilibrium beliefs about the state of the adjacency at date t be correct when averaged over the map's stochastic transition paths. Thus, we have

Definition: Let $F_0^\infty(X)$ (or F_0^∞ in short) be the stationary distribution over the state of the map at the initial adjacency equilibrium. An Adjacency Equilibrium over the transition path is a vector of choice probability functions $\{CCP_{it}(X_{(i1t)}, n_{it})\}_{t=1}^\infty$ and of beliefs $\{\tilde{F}_{it}^*(X_{(i1)t+1}|X_{(i1t)}, n_{it})\}_{t=1}^\infty$ such that: a) CCPs and Beliefs converge to the CCPs and Beliefs of the Adjacency Equilibrium AE_τ. b) Given beliefs $\tilde{F}_{it}^*$, the decision rule CCP_{it}^* is the solution of the plot owner's i's dynamic game "against nature" at every t; and c) beliefs at each t are correct on "average". That is, let $F_t(X_t; \{CCP_s\}_{s=1}^t, F_0^\infty)$ be the joint distribution over the state induced by the primitives, the vector of CCPs from date $t = 1$ to t and the initial steady state distribution of the map F_0^∞. Further, let $F_t(X_{(i2)t}|X_{(i1)t}, n_{it}\{CCP_s\}_{s=1}^t, F_0^\infty)$ be the conditional distribution implied by $F_t(X_t; \{CCP_s\}_{s=1}^t, F_0^\infty)$. Then,

$$\tilde{F}_{it}^*(X_{(i1)t+1} = x_{(i1)t+1}|X_{(i1t)} = x_{(i1)t}, n_{it}) = \sum_{x_{(i2)t}} F_t(x_{(i2)t}|x_{(i1)t}, n_{it}; \{CCP_s\}_{s=1}^t, F_0^\infty)$$
$$F_t(x_{(i1)_{t+1}}|x_{(i1)t}, x_{(i2)t}, n_{it}; CCP_t). \tag{H.1}$$

H.2 Solution algorithm

Recall that in our empirical structural model we futher reduce the dimensionality of the Adjacency Equilibrium by partitioning the set of adjacencies into types, such that all adjacencies of the same type share beliefs and CCPs. Furthermore, the state of the adjacency is reduced to $\mathcal{N}, n$. We compute am Adjacency Equilibrium over a transition path as follows.

Step 0 Solve for the steady state in the benchmark no-tax economy ($\tau = 0$) using the algorithm in the main text and recover the steady state distribution F_0^∞.

Step 1 Solve for the steady state in the counterfactual economy ($\tau > 0$) using the algorithm in the main text and recover the value function for each plot type (V_T)

Step 2 Assume that the village converges to this counterfactual steady state and that it is in this steady state in period T.

Step 3 Guess a sequence of beliefs $\{\tilde{F}_t\}_{t=1}^T$ (as an initial guess, linearly interpolate beliefs from the benchmark to the counterfactual steady state)

Step 4 Solve for plot owner's decision as follows:

Step 4.1 Start in period T-1

Step 4.2 Given value function V_T and beliefs $\tilde{F}_{T-1}$, solve for the CCP_{T-1} and recover V_{T-1}

Step 4.2 Iterate until $t = 1$ and recover $\{CCP_t\}_{t=1}^T$

Step 5 Given $\{CCP_t\}_{t=1}^T$, sample an initial state of the map from the benchmark economy in the steady state F_0^∞ and simulate a transition of well drilling decisions, successes, and failures in every plot on the map from $t = 1$ to T. Replicate this simulation N_S times, e.g. $N_S = 250$.

Step 6 From the transition simulations, construct estimates of the one-period ahead state transition matrices $F_t(\mathcal{N}'|\mathcal{N}, n)$ for each plot type (i.e., averaging across plots on the map of the same type at the same date). Update beliefs and go back to Step 4 and continue iterating steps 4-5-6.

Step 7 Continue iterating until the incremental change in $\{CCP_t\}_{t=1}^T$ is sufficiently small.