Over-drilling: Local Externalities and the Social Cost of Electricity Subsidies in South India*

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

Abstract

Borewells for groundwater extraction have proliferated across India in recent years, partly in response to massive electricity subsidies. Because borewell discharge is attenuated by other wells operating nearby, farmers interact strategically with potentially many neighbors in deciding whether and when to drill. Using survey data from two districts of southern India, we establish both the importance of this well interference externality and its influence on drilling decisions. We then estimate a structural model of well-drilling featuring a dynamic discrete investment game played across a network of adjacent plots. Using the model to compare the social value of groundwater under free (but rationed) electricity against a counterfactual annual charge fully defraying electricity costs, we find that the latter policy, by reining in over-drilling, increases social welfare by 28%. Moreover, the optimal tax on new borewells is equal to electricity costs plus a Pigouvian premium of about 30% to correct the externality.

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

Recent decades have witnessed a groundwater revolution across the Asian subcontinent as millions of borewells have sprung up, 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). Most Indian states, however, 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 a form of rent-seeking ("drilling for subsidies") as smallholders sink costly wells that would not be economically viable absent these price 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). Farmers in these states typically run their pumps continuously during the fixed number of hours (7-9) per day that electricity is available for agricultural use (Fishman et al., 2023). As pumping begins, the water table around each well drops, creating a conical draw-down region centered on the borewell. If two wells are located close enough to each another, their respective cones of depression will overlap, significantly reducing the flow from both wells.[2] 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). During the dry season, when groundwater is virtually the sole source of irrigation, well interference becomes an important, albeit highly localized, common property externality.[3]

To be sure, even in the absence of this externality, electricity subsidies create welfare losses by encouraging borewell over-investment, as the gross return from the marginal well falls below the private cost of drilling plus the fiscal cost of the subsidy. Absent externalities, the social cost of over-drilling could be assessed using a partial equilibrium framework,[4] in which one simulates the number of borewells sunk (per unit area) and their expected private return under a counterfactual policy of removing subsidies.

[1] In 2013, the Indian government 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. 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). 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] Well interference is accentuated in hard rock aquifers, such as those of south India, because of the low transmissivity (velocity of horizontal groundwater flow). Blakeslee et al. (2020) describe the process of well failure in such aquifers in greater detail, emphasizing the importance of *local* hydrogeological features, i.e., sub-surface fractures fed from different sources of recharge.

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

With spatial externalities, 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 strategic interactions between neighboring farmers can generate valid policy counterfactuals.

To appreciate the complexity of our problem, consider the related setting in Lin (2013) that estimates a strategic model of firms exploring and developing (i.e., drilling) off-shore oil parcels in the Gulf of Mexico. To simplify the estimation, Lin restricts the sample to adjacent pairs of (oil) parcels taken in isolation. We cannot use this approach, however, because any given farmer's decision to drill is influenced by nearby wells and by the decisions of neighboring farmers, who in turn are influenced by wells in their neighborhood and by the decisions of all their surrounding neighbors, and so on ad infinitum. Similarly, using (arbitrary) administrative boundaries to demarcate 'neighborhoods' within which all farmer/plots are treated symmetrically (e.g., Fu and Gregory 2019) is problematic in our case because the externality is highly localized.

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 then allows for a novel estimation strategy.[5] 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. This adjacency equilibrium is solved within an estimation algorithm along the lines suggested in Aguirregabiria and Mira (2010).

Using a two-year panel (2010 and 2017) of borewell discharge (or flow) and failure data from 1,057 reference plots, we provide evidence of well interference effects. We also use a 5-year panel (2012–2016) of well-drilling attempts (successful and otherwise) in each of these 1,057 plots, along with panel data econometric methods that address endogeneity of past investments and recall/measurement error, to show that drilling is less likely the more functioning borewells there are on the plot *and* in the adjacency *outside* of the plot. Taken together, these results establish both the importance of the local externality and its influence on farmers' investment decisions.

With the 'first-stage' results for well flow and failure processes as inputs, along with subjective

[5] In our sample of plots described below, the average number of functioning borewells remains relatively stable at around two over a 5-year period despite a roughly 7 percent annual failure rate.

drilling success rates, information on drilling costs, and cadastral maps showing plot configurations, we pin down primitive parameters of the production technology by estimating the dynamic structural model via simulated method of moments (SMM). In particular, we match, among other moments, the observed drilling rates by plot size and by the number of currently functional borewells on the plot to the corresponding conditional choice probabilities. Given reduced-form evidence that well-drilling is strongly increasing in farmer wealth, we also assume that investment can only occur on plots owned by farmers with sufficient wealth.[6] To identify the sample proportion of such 'developed' plots we target additional moments related to drilling rates by wealth level.

Using the estimated structural parameters, we find in the steady state that the expected discounted present value of agricultural output minus drilling costs on developed land is 45,600 Rs/acre. By contrast, the value to *society*, which differs from the private value by the expected discounted present value of electricity costs, is only 14,500 Rs/acre. Moreover, charging borewell owners fully for electricity through an annual flat rate would increase the social value of groundwater development to 19,000 Rs/acre, or by 28%. In other words, the social cost of free electricity amounts to 4,500 Rs (65 US$) per acre, or around 1.5% of farmland value in our setting.

Finally, we consider the optimal borewell tax that both eliminates rent-seeking due to free electricity and the externality costs due to well interference. 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. We find that, when both new and old borewells are subject to taxation, the social welfare maximizing tax equals annual electricity costs plus a Pigouvian premium of around 12% to correct for the externality. When existing borewells are grandfathered, however, the social welfare maximizing tax equals annual electricity costs plus a Pigouvian premium of 30% (about 2,500 Rs per borewell per year); the higher premium is required because the marginal externality cost of a new borewell is higher when old borewells still operate. Nevertheless, both tax policies curb excessive drilling and thus achieve similar social welfare along their respective transition paths. Grandfathering of existing borewells is, however, the more politically palatable policy as no farmer would suffer a capital loss on sunk investments.

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 consisted entirely of static applications (see, e.g., Acemoglu et al. 2015, Xu 2018, König et al. 2017).

[6] For structural models of investment incorporating liquidity constraints in similar contexts see Rosenzweig and Wolpin (1993) and Fafchamps and Pender (1997).

While we could have also used a static model in which all plot owners drill their borewells simultaneously, such a model would not account for well failure, an inherently dynamic concept featuring prominently in our setting. Moreover, a static model would not inform us about the transition dynamics between the baseline scenario of over-drilling to a new steady state under alternative counterfactual policy regimes. Analysis of transition paths is crucial to distinguishing between different policy treatments of existing wells (see, e.g., Domeij and Heathcote 2004 in the context of capital taxation).

We also contribute to a small empirical literature quantifying the social costs of common property externalities. Closest in spirit to our work is the aforementioned oil-drilling paper by Lin (2013). In addition to limiting its analysis to pairs of interacting parcels, Lin (2013)'s approach does not accommodate another important feature of our setting, permanent unobserved heterogeneity in the suitability of land for resource development. 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 by one boat reduces the potential catch of all other boats equally. In the context of 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. Finally, Ryan and Sudarshan (2022) estimates the welfare cost associated with over-pumping in the north Indian state of Rajasthan, without considering well interference, focusing instead on the non-local, aquifer-wide, externality.[7] Ryan and Sudarshan (2022) also takes the number of borewells as given, thus ignoring drilling costs which are 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 in this paper.

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 analysis of counterfactual policies and the optimal borewell tax follow in Section 6. Section 7 concludes.

[7]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 such complications in our study, we deliberately chose a setting in which groundwater markets are virtually nonexistent (see Giné and Jacoby 2020).

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 power, a move swiftly followed by the major states of Tamil Nadu, Karnataka, and Punjab.[8]

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'.[9] 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.[10] Well interference, therefore, is the predominant groundwater pumping externality in our setting, one that is both localized and *static*, affecting only current groundwater availability.

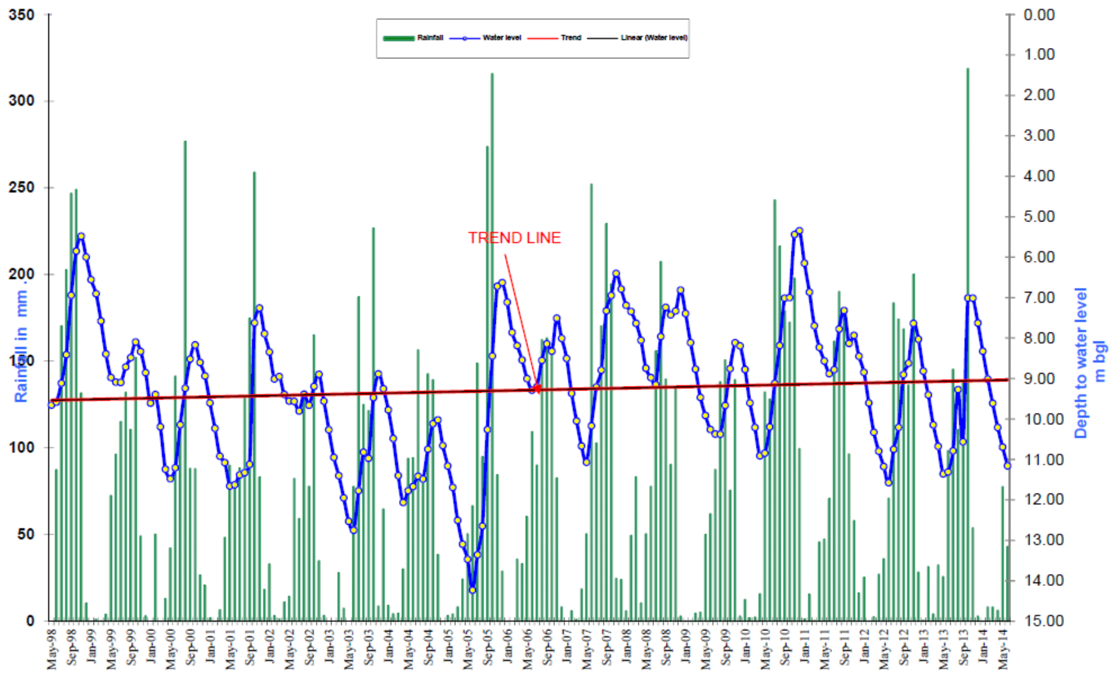
Our data come from the drought-prone districts of Anantapur (Andhra Pradesh) and Mahabubnagar (Telangana), originally the backdrop for the weather insurance study of Cole et al. (2013). As shown in Giné and Jacoby (2020), groundwater availability and the related development of groundwater markets in 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

[8]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 USD 450 for the entire state of AP circa 2010, whereas we obtain a much more conservative figure of USD 180 (8500 Rs) in our study areas (see Appendix A).

[9]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 and Ryan and Sudarshan 2022).

[10]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 C 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).

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.

An important contextual feature contributing to well interference is the high degree of land fragmentation. Indeed, absent such fragmentation, the well interference externality would be internalized through unified landownership. To replicate the 'networks' of individually owned land parcels across which to compute the drilling-game equilibria for our structural estimation, we manually digitized cadastral maps for a subset of study villages, at least one for each mandal or county (see Appendix D for details and an example of the original cadastral map and its digitized counterpart). In all, we have 14 maps containing 12,330 land parcels with a median size of 2.02 acres.

2.2 Adjacency panel: Well drilling

Our main dataset is a seven-year (2011-17) retrospective panel on borewell status and drilling attempts. In 2017, we were able to re-interview 1,436 of 1,488 farm households that participated in the the 2010 survey used in the weather insurance study of Cole et al. (2013). The 2017 survey instrument included a history of drilling attempts on and around each of the household's plots since 2011. Based on these responses, we selected up to one reference plot per respondent if there had been drilling attempts made in the last seven years either on the plot itself or within a 500 meters radius of the plot (in case of two or more eligible plots per household, one was chosen at random).

An *adjacency* is a central concept underlying both our data collection and our structural model; it is defined as the set of all agricultural plots contiguous to the reference plot, inclusive of it. As part of the 2017 instrument, we administered an adjacency survey to the 1,057 farmers with an eligible reference plot, asking the 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. We shall assume throughout this paper that only other borewells operating in the adjacency create interference effects on the reference plot, imposing a negative externality. Put differently, the effects of borewells outside the adjacency on reference plot borewell flow and failure are negligible. Given the typical size of plots and the range of well interference effects in our setting (Chandrakanth, 2015), we believe that this is a sensible assumption.[11]

Denoting by n_{it} the number of functioning wells on reference plot i in year t and by $\mathcal{N}_{it}$ the number of functioning wells in the adjacency *outside* of reference plot i, we define the total number of functioning wells in the adjacency as

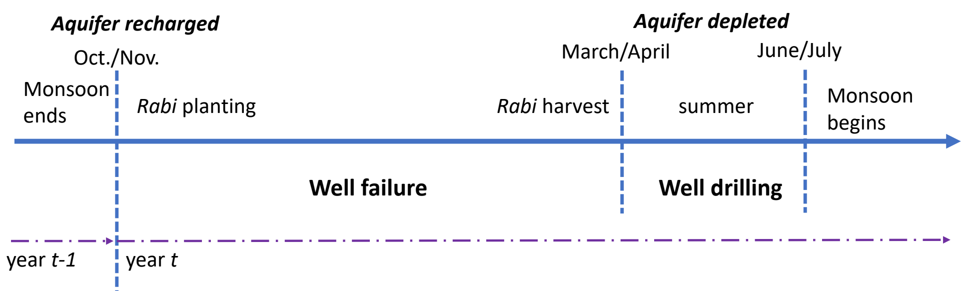
$$N_{it} \equiv \mathcal{N}_{it} + n_{it}. \tag{1}$$

Arguably, respondents may less accurately recall the status of borewells on adjacent plots than of those on (their own) reference plots. We, therefore, allow $\mathcal{N}_{it}$, but not n_{it}, to be measured with error, specifically, misclassified as functioning or failed. We discuss the econometric implications of this form of measurement error in subsection 3.2.

Figure 2 provides an event timeline, with the "year" beginning at rabi season planting. 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 $t+1$. Well failures can occur throughout the rabi season; our survey does not record the exact month of failure. We, therefore, treat year t failures in the same way as successful drilling attempts in year t, assuming that neither event

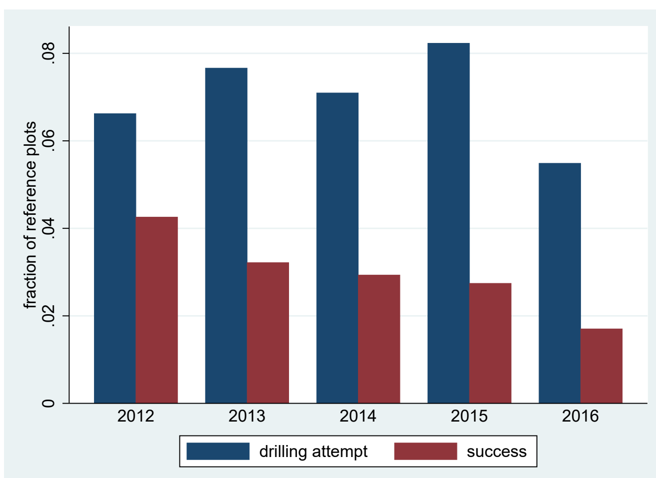
[11]The median reference plot in our data is about one hectare. In a chessboard configuration of identical plots of this size with borewells located at the center, the distance between a borewell in the reference plot and one elsewhere in the adjacency would be 100-140 meters, well within the range of 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.

Figure 2: Timeline



affects agricultural output or the decision to drill in year t.[12] In other words, year t drilling, output, as well as failure depend on the number of functioning borewells at the end of year $t-1$.

Figure 3: Drilling attempts and successes by year



We drop the data from 2011 because we do not have the lagged (2010) number of functioning wells for 2011. We also drop the data from 2017 because, as the survey was administered in May of 2017, well drilling attempts and failures that may have occurred later in 2017 were not recorded, potentially

[12]Well failure is also not necessarily seen as a discrete event. If a well fails to yield water at the end of the rabi season, for example, it may only be at the beginning of the next rabi that it is considered to have truly failed.

understating drilling activity for that year. From 2012-16, a total of 371 reference plot drilling attempts were reported in $1,057 \times 5$ adjacency-years, for a 7.0% average annual rate of drilling (see Figure 3); 157 or 42.0% of these attempts were successful.

Lastly, we note that, over 2012-16, the average number of own borewells n_{it} per reference plot (0.648) is substantially greater than the average of $\mathcal{N}_{it}$ per adjacent plot (0.385). Given our sampling procedure described above, we conjecture that this difference in well density reflects selection on the basis of factors, idiosyncratic to the plot or to its owner, that make the reference plot particularly conducive to drilling on, a point that we shall return to later in regards to the structural estimation.

2.3 Reference well panel: Failure

To estimate the annual probability of well failure, we construct a 2012-16 panel of reference plot borewells that are *at risk* of failure. For reasons that will become clear below, when the reference plot has multiple functioning wells we restrict attention to the oldest, i.e., the first one sunk. We also drop wells sunk in 2016 which would not have had time to fail; more generally, wells only enter the failure panel in the year after they are sunk. Since failure is an absorbing state, a well exits the panel in the year following its failure. The result is an unbalanced failure panel of 697 borewells over as many as five years. This figure of 697 is lower than the total number of 1,057 adjacencies because 320 adjacencies have no functioning borewells on the reference plot over the 2012-16 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 B.1).[13]

A key issue in modelling well failure is duration dependence, whereby the probability of failure depends on the age of the well. If water tables are 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, 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 the 606 extant borewells in 2012, when they had a median age of 12 years. Regressions reported in Appendix C reveal a significant positive association between well age and subsequent failure, but this entire effect is driven by 59 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.

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

2.4 Well flow

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.

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 B.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 B.1) rainfall during the 2016 monsoon season, responsible for rabi 2016-17 recharge, fell roughly 30% short of that in 2009.

2.5 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) is that they provide a *probability*, albeit a subjective one, around each plot regardless of whether there was ever an attempt to drill on it. Because subjective assessments of drilling success could be affected by recent drilling activity, however, we only use responses from the 2010 survey, predating our drilling panel (see Appendix Figure B.2). While there is considerable variation in the average subjective success rate across mandal (ranging from 0.337 to 0.572), the overall average rate of 0.446 is remarkably close to the actual success rate of 0.423 for 2012-16 noted above.[14] 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 model assumes that the drilling success probability, proxied by the mandal average subjective rate, is independent of both the number of functioning wells in the adjacency and of monsoon rainfall.

[14] This relatively low success rate reflects the nature of the shallow hard rock aquifer, where groundwater storage is highly localized, confined to certain narrow fissures. By contrast, drilling success rates in a deep alluvial ("bathtub"-type) aquifers would be close to one.

3 Preliminary Estimation

In this section, we first develop the panel data econometrics for our preliminary estimations using the survey data described in the previous section, then investigate 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 that we shall use as primitives in the structural estimation.

3.1 Unobserved heterogeneity

To allow for time invariant unobserved heterogeneity, we specify the probability of an outcome as a function of an index in plot i and period (year) t as follows:

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

where ν_i is a normally distributed random effect that varies across reference plots, N_{it-1} is the number of functioning wells in the adjacency at the end of the previous period, and R_{mt} is a dummy variable that takes a value of one if monsoon rainfall in mandal m in year t 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-1}, R_{mt}). 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-1} and R_{mt} being strictly exogenous conditional on ν_i.[15] In particular, let

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

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

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

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

[15] In other words, N_{it-1} and R_{mt} must be uncorrelated with all present and future values of ε_{it}. In the case of the reduced-form drilling logit, however, strict exogeneity is violated if we condition on $N_{it-1} = \mathcal{N}_{it-1} + n_{it-1}$ (or $\mathcal{N}_{it-1}$ and n_{it-1} separately) because a successful drilling attempt in period t will augment n_{it} by one. For this reason, we adopt a slightly different approach in the reduced-form drilling logit discussed in Subsection 3.3.

3.2 Misclassification error

The reported number of functioning wells on *neighboring* plots, $\mathcal{N}_{it-1}$, is plausibly subject to recall error, as already noted, leading to a non-classical form of measurement error.[16] We thus begin with the (also plausible) assumption that the number of *existing* neighboring wells $\mathcal{N}^E_{it-1}$ is accurately observed. Ignoring n_{it-1}, 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 $Pr(Y_{it}|\mathcal{N}_{it-1})$. Although $\mathcal{N}_{it-1}$ is not perfectly observed, we know that

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

where $Pr(k|\mathcal{N}^E_{it-1})$ is the discrete probability density of the true number of functioning wells outside of the reference plot. This density is of the binomial form

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

where p is the underlying annual probability of well failure. For simplicity, we take p to be constant over time and across adjacencies in each of the two districts. From our data on the rate of reference plot borewell failures, we estimate $\hat{p} = 0.104$ in Anantapur and $\hat{p} = 0.052$ in Mahabubnagar.

Our misclassification error model (MEM) estimator then assumes that the likelihood contribution conditional on unobservable μ is

$$\ell^y_i(\mu) = \prod_{t=1}^{T} \sum_{k=1}^{\mathcal{N}^E_{it-1}} Pr(Y_{it}|z_{it}(k,\mu))Pr(k|\mathcal{N}^E_{it-1}, \hat{p}). \tag{7}$$

3.3 Determinants of drilling

Our drilling logit is based on CRE-MEM likelihood (7) with μ assumed normally distributed. A successful drilling attempt in period t, however, increases the number of functioning wells on the reference n_{it} by one well. Thus, we have in n_{it-1} essentially a lagged endogenous variable. Wooldridge (2005)

[16]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.

suggests controlling for the initial conditions, altering equation (4) as follows:

$$\begin{aligned} z_{it}^d &= \mu_i + \beta_0 + \beta_1 \mathcal{N}_{it-1} + \beta_2 \mathbb{1}_{n_{it-1}=1} + \beta_3 \mathbb{1}_{n_{it-1}>1} + \beta_4 R_{mt} \\ &+ \gamma_1 \bar{\mathcal{N}}_i + \gamma_2 \mathbb{1}_{n_{a0}=1} + \gamma_3 \mathbb{1}_{n_{a0}>1} + \gamma_4 \bar{R}_m + \gamma_5' Z_i + \varepsilon_{it}. \end{aligned} \tag{8}$$

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.

The premise of our theoretical framework, laid out in the next section, is that well interference induces strategic substitutability between drilling decisions of neighboring plot owners, implying that $\beta_1 < 0$. Since $\mathcal{N}_{it-1}$ 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] Turning to the estimation results in Table 1, the column 1 specification, which ignores misclassification of neighboring wells functioning status, shows no significant effect on drilling (first row). The corresponding MEM estimate in column 2, which corrects for misclassification, is much larger in magnitude and statistically significant, indicating that having more functioning wells in the adjacency, outside of the reference plot, reduces the likelihood of drilling.[18] We thus find strong evidence of strategic substitutability between neighbors' drilling decisions.

Also in Table 1, we see that having more functional borewells on the reference plot substantially reduces well-drilling by the plot owner, perhaps due to diminishing returns to groundwater in production and to potential interference among own borewells.[19] As expected if rainfall were iid across years and thus not predictive of future rainfall, a good past monsoon does not significantly affect the propensity to drill on the reference plot. We also find more drilling on larger plots. However, drilling is unaffected by the number of plots in the adjacency, which supports a simplified state-space (N, n) for the reference plot owner's investment decision (see Section 4.2 below). All of these results, and that on strategic substitutability, are robust to the inclusion of mandal dummies (compare columns 2 and 3).

[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. Of course, if the random effect μ_i fully accounts for serial dependence in the error term of equation (8), then the CRE-MEM estimator of β_1 is unbiased regardless of spatial correlation.

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

[19]The marginal effects are very large: Going from 0 to 1 reference plot borewells reduces the predicted annual drilling rate from 0.132 to 0.044, whereas going from 1 to 2 reference plot borewells reduces the predicted annual drilling rate from 0.044 to just 0.017.

Table 1: Determinants of Drilling

	(1)	(2)	(3)	(4)
No. func. wells outside ref. plot ($\mathcal{N}$)	-0.143	-0.549	-0.599	-0.626
	(0.170)	(0.211)	(0.203)	(0.205)
1 func. well on ref. plot ($\mathbb{1}_{n=1}$)	-1.083	-1.150	-1.092	-1.124
	(0.195)	(0.199)	(0.195)	(0.195)
2+ func. wells on ref. plot ($\mathbb{1}_{n=2}$)	-2.204	-2.350	-2.065	-2.103
	(0.373)	(0.383)	(0.375)	(0.376)
Good monsoon (R)	-0.0622	-0.0494	-0.0608	-0.0616
	(0.171)	(0.172)	(0.186)	(0.186)
Log(ref plot area)	0.366	0.372	0.337	0.192
	(0.0850)	(0.0858)	(0.0883)	(0.0947)
No. of plots in adjacency	-0.0332	-0.0381	0.0170	0.0251
	(0.0631)	(0.0636)	(0.0629)	(0.0629)
Log(gross wealth in 2009)	—	—	—	0.284
				(0.0787)
Mandal dummies (11)	No	No	Yes	Yes
Estimation method	CRE	CRE-MEM	CRE-MEM	CRE-MEM
Log-likelihood	-1296	-1293	-1271	-1265
H_0 : No plot-level unobserved heterogeneity				
LR test statistic	68.11	148.5	93.79	86.58
p-value	0.000	0.000	0.000	0.000

Notes: Standard errors in parentheses. Dependent variable is an indicator for whether well-drilling was attempted on reference plot that year. All estimations use a sample of 1,057 reference plots over five years (for a total sample of 5,285). All logit models estimated by maximum likelihood (selected coefficients reported). For estimation method, CRE refers to correlated (normally distributed) random effects (see subsection 4.1) and MEM to misclassification error model (see subsection 4.2).

Next, we assess the importance of financial liquidity for borewell investment. From the 2010 household survey, we construct a measure of the reference 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.[20] Not surprisingly, given the high capital requirement of drilling a borewell, initially wealthier households were significantly more likely to undertake such investment between 2012-16 (column 5). In particular, a doubling of financial wealth would increase the predicted annual drilling rate from 7.2 to 9.5%. Since households with larger reference plots also tend to be wealthier, conditioning on wealth attenuates the reference plot area coefficient by more than 40% (compare columns 4 and 5), although it remains statistically significant.

Lastly, for each specification of Table 1, we provide a likelihood ratio test against the null of no unob-

[20]In this low-income agrarian setting, it is very difficult to distinguish between liquid and illiquid assets and we make no attempt to do so. Total wealth should be a good proxy for liquidity.

served 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 all specifications, we strongly reject the null of no heterogeneity.[21] 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, therefore, 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.[22] 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 $c_6 = \infty$. Note that the variable $\mathcal{N}_{it}$ differs between the household/plot panel dataset used in the flow estimation and the retrospective adjacency panel dataset. In the former case, farmers were asked about the number of other functioning borewells within a 100 meters radius of the reference plot; while not precisely the same as our concept of adjacency, the two definitions of neighborhood turn out to be essentially the same in practice (see below). Moreover, since the $\mathcal{N}_{it}$ obtained from the household/plot survey is the contemporaneous (rather than retrospective) report of the respondent, we assume no misclassification error.

Failure For reasons discussed in Section 2.2, we adopt a constant failure hazard specification, using the sequential logit as in Cameron and Heckman (1998), among others. The conditional likelihood

[21] 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 reported in the table.

[22] 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.

contribution with $\mathcal{N}_{it-1}$ subject to misclassification error, as discussed in Subsection 3.2,[23] is

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

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 also noted in Section 2.2, the failure panel is restricted to just one reference plot borewell; we always choose the oldest well. Allowing multiple borewells would lead to a violation of strict exogeneity due to correlation between N_{it-1} 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.[24]

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] Whether derived from the adjacency or household/plot survey (and conditional on having at least one functional reference plot borewell), N_{it-1} averages very close to 3 borewells in 2017, indicating that an 'adjacency' and a 100 meter radius around a plot are essentially the same thing.

[24] 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).

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 $H_0 : \sigma_\nu = \kappa_\nu = 0$.[25] 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 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 controls for misclassification error using the MEM approach; and column 3 augments the column 2 specification with mandal dummies in both flow and failure linear indices. Consistent with the well interference externality, having more borewells in an adjacency depresses well-flow and makes failure of the reference well more likely. This latter effect, however, only emerges when we control for misclassification error using MEM 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 column 3 shrinks the estimated standard error 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).

Moving to Step 4, we redo Step 2 allowing for 2 *discrete* types, obtaining a log-likelihood value of -2245.59 (compare to -2246.55 in column 3 of Table 2). Since adding a third type does not lead to an 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).

Note that the flow-failure estimation sample is restricted to reference plots that had a functioning borewell at some point during 2010–2017, which we shall refer to as 'developed' plots. Our four-step procedure recovers the distribution of unobserved type $\nu = \nu_f$ only for such developed plots. In the

[25] 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.

Table 2: Joint Flow-Failure CRE Estimation

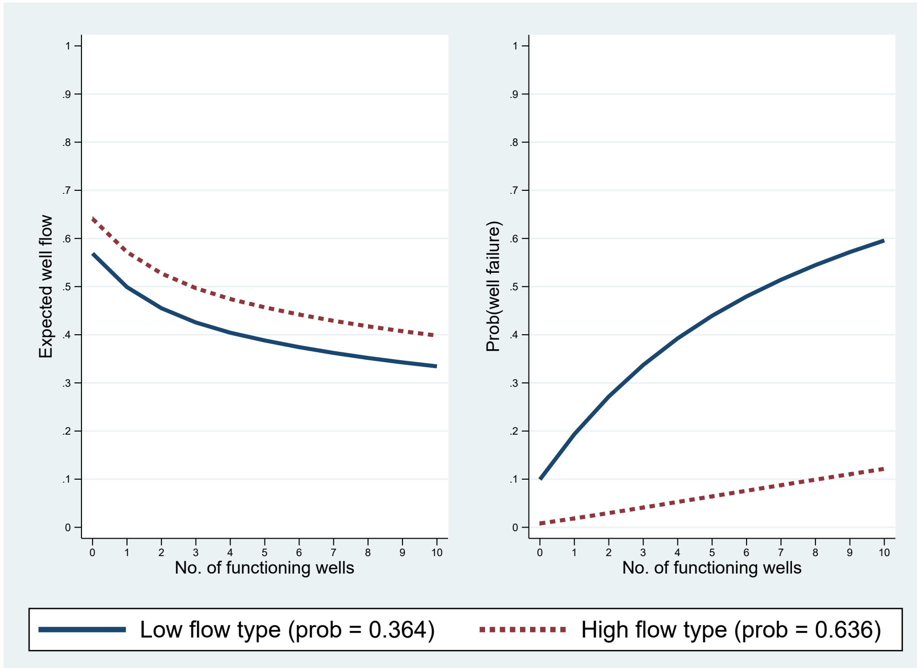
Step 1	(1)	(2)	(3)
Flow:			
log(N)	-0.849	-0.858	-0.884
	(0.171)	(0.172)	(0.172)
Good monsoon	1.766	1.782	1.917
	(0.800)	(0.802)	(0.804)
Failure:			
log(N)	0.052	1.842	1.163
	(0.635)	(0.560)	(0.473)
Good monsoon	0.119	0.127	-0.258
	(0.211)	(0.222)	(0.259)
Mandal dummies	NO	NO	YES
Estimation method	CRE	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	-2404.41	-2480.97	-2246.55
H_0 : No plot-level unobserved heterogeneity			
LR test statistic	123.203	218.726	30.789
p-value	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.

structural estimation, we mimic this approach by drawing an unobserved plot type from the estimated distribution *only* if the plot could be developed (see Subsection 5.2 for details).

To summarize the empirical results thus far, more functional borewells near a plot reduce the discharge of borewells on that plot, increase their likelihood of failure, and reduce the propensity for further drilling. Taken together, these findings point to a well interference externality that farmers incorporate into their investment decisions. We turn next to the theoretical framework for these decisions.

Figure 4: Expected flow and failure probability by type



4 A Model of Borewell Investment

4.1 Preliminaries

Let the *incremental* agricultural output from having 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 production function parameters, a_i is plot area, and q_{it} is well discharge, which is stochastic and thus unknown to the farmer prior to drilling. We assume that borewell discharge has a discrete distribution with K points of support $\{q_{it1}, ..., q_{itK}\}$ each with probability π_{itk}. Along with a constant elasticity of substitution (CES), production function (13) also imposes constant returns to scale (CRS), so that output per acre only depends on total borewell discharge per acre.[26] The scale parameter θ converts physical units of incremental output into 2017 Indian Rupees (Rs).

Discharge probabilities π_{itk} evolve over time depending on monsoon rainfall, R_t, which determines

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

annual aquifer recharge, and on the number of functioning borewells in the adjacency N_{it} according to

$$\pi_{itk} = \pi_k(N_{it}, R_t). \tag{14}$$

The well interference externality, estimated in Subsection 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 $\pi_{Fit} = \pi_F(N_{it}, R_t)$. Once a borewell fails, it will never have a positive discharge again regardless of monsoon rainfall R_t, and, as a result, $y_{it} = 0$ forever.[27] Drilling success is also stochastic and, for reasons discussed in Subsection 2.5, we assume that the probability of success π_S is constant.

Drilling a borewell entails a cost c_d and if the attempt is successful, there is an additional cost of installing a pipe, casing, and applying for the electricity connection (the submersible pump itself is removable and thus not considered a sunk cost). The total cost of a successful attempt is thus $c_s > c_d$ and both c_s and c_d are estimable from our survey data.

Finally, for the sake of tractability, we assume that at most two wells can function simultaneously on any given plot.[28] So, if p_i is the number of plots in adjacency i, then $N_{it} \in \{0, .., 2p_i\}$. 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 Subsection 3.1). 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

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

[27] 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 a permanent (i.e., 'absorbing') state, which is indeed what we find in Subsection 3.4.

[28] In our sample, there are just 21 out of 5,285 plot-years in which a reference plot had 3 functioning wells at the same time (zero cases of 4 or more wells). We set $n_{it} = 2$ in these cases.

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 $CCP(\mathcal{N}, n) \equiv \Pr(d = 1 \mid \mathcal{N}, n)$.[29]

We first describe the dynamic investment problem 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 is assumed to consist only of the number of own functioning wells, $n \in \{0, 1, 2\}$, and the total number of wells in other plots in the adjacency, $\mathcal{N} \in \{0, .., 2(p-1)\}$. In the next subsection 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, are iid across choices and time, and are 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, we have

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

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 $\tilde{F}(\mathcal{N}' \mid \mathcal{N}, n)$ captures the individual's 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 this period and on n functioning

[29] While subscripts are removed for expositional ease, according to the timing conventions of Section 2.2, the drilling decision on reference plot i at time t, d_{it}, depends on n_{it-1} and $\mathcal{N}_{it-1}$.

wells on the reference plot this period ($n = 0$, in this case). As indicated in Figure 2, since drilling occurs only after (rabi) production has taken place, the increase in expected output from any successful attempt are 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

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

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

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

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. As previously described, we use 14 cadastral maps representing at least one village in each mandal. While the borders of these administrative maps are arbitrary in that they do not correspond to salient geographic or geological features, since each contains many plots, "truncation-at-border" effects should have negligible empirical consequences.[30]

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 the plot's unobserved 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.

The state of plot i in period t is defined by the number of functioning wells $n_{it} \in \{0, 1, 2\}$. 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. In other words, under *unrestricted* MPE play, investment decisions may depend on the state of the whole map, even with well interference effects confined to adjacent plots. 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"; and b) beliefs are correct, in that $\widetilde{F}^*(X_{t+1} \mid X_t) = F(X_{t+1} \mid X_t; \{CCP\}^*)$.

[30] 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.

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

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, firstly, 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.[31] Thus, in the spirit of an "oblivious equilibrium",[32] 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.[33] Further, let $F^\infty(X_{(i2)t} \mid X_{(i1)t}, n_{(it)}; \{CCP\})$ be the conditional distribution implied by $F^\infty(X_t; \{CCP\})$. Then,

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

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, i.e., the states of plot j and those of all its neighbors. The neighbors of plot j are plot i and some (but not necessarily all) of the other plots 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 $X_{(i2)t}$,

[31]Even if all plot owners base their drilling decisions solely on the state of their own adjacencies, the evolution of the state of any adjacency more than one period ahead will still depend on the state of the full map today. There is, therefore, a tension between reducing the state space of the decision rule and imposing the equilibrium constraint of coherency between beliefs and behavior. This tension is resolved by imposing coherency in steady state.

[32]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 reseach.

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

he would thus be able to predict his neighbor j's behavior at t using $CCP_j^*()$ and, together with other primitives such as the drilling success and well failure processes, predict the stochastic evolution of the state of plot j; this prediction is the second factor in each term of the summation in equation (23). An AE assumes, however, that, rather than having $X_{(i2)t}$ in his information set, plot owner i can only form expectations about it using the (conditional) steady state distribution $F^{\infty}(X_{(i2)t} \mid X_{(i1)t}, n_{it}; \{CCP\}^*)$, hence the probability weights on the RHS of equation (23).[34] Although plot owners each still have a unique CCP and beliefs, and their joint decisions still depend on the entire village map, the AE concept achieves considerable simplification.[35]

5 Structural Estimation

5.1 Overview

Three results established in Sections 2 and 3 inform our empirical strategy: First, misclassification error in the reported number of functioning borewells on neighboring plots leads to non-trivial econometric biases. Since incorporating a misclassification error correction into a likelihood-based estimator is complex, we pursue the simpler SMM approach described below. Second, there is important plot-level unobserved heterogeneity in groundwater availability driving both well flow and failure, which we shall have to account for in our estimation algorithm. Third, the adjacency panel, which we shall use for the structural estimation, is restricted to plots from the Cole et al. (2013) weather insurance study that had any drilling activity within a 500 meter radius over the previous seven years; this drilling activity requirement was satisfied by (only) 74% of surveyed farmers. Furthermore, we have evidence that lack of drilling activity may reflect liquidity constraints, as drilling is significantly positively associated with the plot owner's initial wealth conditional on the number of currently functioning borewells on that plot. We discuss how we tackle selection and liquidity constraints in the next subsection.

Before laying out our estimation algorithm, we summarize the primitives of the structural model in Table 3. In addition to the four structural parameters $\Omega = (\theta, \alpha, \delta, \sigma)$ that we estimate in the second stage, the estimation algorithm recovers three auxiliary parameters, to which we now turn.

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

[35] 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
Success probability (mandal-level):	π_S	2.5
Successful drilling cost	c_s	note 3
Unsuccessful drilling cost	c_d	note 4
Good monsoon prob. (mandal-level):	$\mathbb{E}\, R_{mt}$	3.1
Fixed parameters:		
Discount factor	β	note 5
Map of plot network (mandal-level)	$\{A, \mathbf{M}\}$	4.3

Notes: See subsection specified in column 3 and/or the following numbered notes for details: (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) $c_d = 35,200$ Rs. is computed as average drilling cost (in 2017 Rs) across all borewells sunk since the year 2000; (4) $c_s = 72,300$ Rs. is computed as c_d plus the average cost of the pipe, casing, and electrical connection across all borewells sunk since the year 2000; (5) $\beta = 0.90$ throughout estimation.

5.2 Empirical specification

Selection and liquidity constraints: To adequately fit the observed distribution of drilling and borewells across reference plots in our adjacency sample to the model's predictions for a plot network (map) of an entire village, we assume that a fraction of plots are *permanently* incapable of being developed for groundwater extraction; i.e., for these plots, $n_{it} = 0$ forever and the annual decision about whether to drill or not to drill is irrelevant. Let us refer to these plots as *undeveloped* and to all other plots susceptible to drilling as *developed.* We (econometricians), however, do not observe which plots with $n_{it} = 0$ are (permanently) undeveloped versus subject to drilling but not currently having a functioning borewell. Let the propensity for a plot to be developed be given by linear index

$$\Lambda_i = \lambda_0 + \lambda_1 w_i + \zeta_i, \tag{24}$$

where w_i is the log of pre-sample wealth of the plot owner and ζ_i is an iid logistic error term. While w_i in equation (24) captures liquidity constraints – that low wealth farmers are less able to afford the drilling costs – ζ_i reflects the (unobserved) suitability of the plot for groundwater development. For the sake of tractability, we shall assume, that ζ_i is uncorrelated with unobserved flow/failure heterogeneity ν_i.[36] We may now write the probability that a plot is undeveloped as

$$\Pr(U_i = 1 | w_i, i \in \text{sample}) = \frac{1}{1 + e^{-(\lambda_0 + \lambda_1 w_i)}}. \tag{25}$$

As described in the next subsection, the initial step of our estimation algorithm assigns a development status, $U_j = 0$ or 1, to each plot j on the map. In doing so, two problems arise: i) the wealth of plot owners on the map is not observed, and ii) the sample of reference plot owners, for whom we do observe wealth, is positively selected on wealth, since drilling is more likely to have occurred on reference plots (see Subsection 2.2) and drilling is positively correlated with wealth through a plot's unobserved development status U. Moreover, plot area a_k (discretized into 4 types indexed by k), which we do observe on the map, is positively correlated with wealth. Consider, then, the probability that a plot on the map of area a_k is undeveloped:

$$\begin{aligned} Pr(U_j = 1 | a_k, j \in \text{map}) &= \int Pr(U_j = 1 | a_k, w_j, j \in \text{map}) Pr(w_j | a_k, j \in \text{map}) dw_j \\ &= \int Pr(U_i = 1 | w_i, i \in \text{sample}) Pr(w_i - s | a_k, i \in \text{sample}) dw_i \end{aligned} \tag{26}$$

where the unknown parameter s in the second line shifts the log-wealth distribution in the sample to the left to mimic the log-wealth distribution on the map. Because of selection on observable wealth, the fraction of undeveloped plots on the map is higher than in the sample and, consequently, there are fewer borewells on average on the map than on our sample reference plots.

State space restrictions and heterogeneity of developed plots: To make the empirical model more tractable, we assume that the CCP in the Adjacency Equilibrium depends on the area of the reference plot a and on the number, but not 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 $(\mathcal{N}_{it}, n_{it})$.[37] We allow for L "types" of developed plots, where type encompasses characteristics that are both observed (area, number of adjacent plots) and unobserved (low or high flow/failure heterogeneity) to the econometrician. The discretization of

[36] Once we obtain estimates of all of the structural model parameters, we shall use simulation to quantify the difference in the distributions of ν_i conditional on the plot being developed and undeveloped.

[37] Our empirical investigation in Section 3.3 indicates that, conditional on $(\mathcal{N}_{it}, n_{it})$, the distribution of functioning wells across plots in the adjacency is not predictive of drilling decisions. Hence, the more fine-grained adjacency state space $(X_{(i1)t}, n_{it})$ would not improve model fit.

reference plot area into four quartiles, already mentioned, coupled with the number of adjacent plots in the maps ranging from 1 to 7, yields 28 possible observed types, along with 2 unobserved types; thus, $L = 56$. Finally, we assume that unobserved plot type ν_i and development suitability ζ_i are both iid across plots.[38]

5.3 Solution Algorithm

Given values of parameters $(\Omega, \lambda_0, \lambda_1, s)$, we obtain an AE on each of the 14 village maps as follows:

Initialize the maps:

Step 1 Given parameter s, randomly draw for each plot j a $\tilde{w}_i$ from the conditional distribution of wealth in the sample $Pr(w_i|a_k, i \in \text{sample})$ and set $w_j = \tilde{w}_i - s$.

Step 2 Given parameters (λ_0, λ_1), draw a U_j from the binomial distribution with $\Pr(U_j = 1|w_j) = [1 + e^{-(\lambda_0+\lambda_1 w_j)}]^{-1}$.

Step 3 Assign each plot for which $U_j = 0$ an unobserved flow type ν_1 or ν_2, drawing from a binomial distribution with probability of (low) type $1 = 0.364$.

Step 4 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 5 Given $\{CCP_{l,k-1} : l = 1, \ldots, L\}$ at iteration $k = 1, 2, ...$, 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 6 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 $\widehat{F}_{lk}$.

Step 7 Given beliefs $\widehat{F}_{lk}$ 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 $\{CCP_{l,k}\}$ satisfying the fixed point condition $CCP_{lk} = \Psi(CCP_{lk-1}, \widehat{F}_{lk}, \Omega)$ for all types, where $\Psi()$ is a policy iteration operator.

Convergence:

[38] In other words, we abstract from spatial correlation in these plot characteristics. Incorporating spatial correlation into the structural model is complex and is left as a topic for future research.

Step 9 If $\|CCP_k - CCP_{k-1}\|$ is small enough, done. If not, update k and back to Step 5. If CCPs converge, so do beliefs which are a continuous function of CCPs.

Steps 1-9 are nested within a routine for minimizing, with respect to parameter vector $(\Omega, \lambda_0, \lambda_1, s)$, the SMM criterion function, as described next.

5.4 Moment conditions and identification

We match the following 22 empirical moments to their model-based counterparts:

1. Probability of drilling for $n = 0$ and $n = 1$ across area quartiles (8 moments);

2. Fraction of plots with $n = 0$ (1 moment);

3. Probability of drilling when $n = 0$ across wealth terciles and area quartiles (12 moments);

4. Fraction of plots with $\mathcal{N} = 0$ (1 moment).

Empirical moments are computed as (cell) means across all reference plot-years, while the weighting matrix used to form the SMM criterion is diagonal, consisting of the inverse moment variances. Heuristically speaking, the first set of moment conditions identify the structural primitive parameters Ω, the second and third set of moment conditions identify the developed plot probability parameters (λ_0, λ_1), and the fourth moment condition (combined with the second) identifies the selection parameter s.

Computing model-based moments corresponding to each of these empirical moments involves simulating drilling decisions and the number of functioning wells on each map for 500 periods in steady state and taking the appropriate averages weighted by the proportion of sample plots linked to each map. This procedure is complicated, however, by unobserved heterogeneity ν and by the presence of undeveloped plots. To fix ideas, consider the model-simulated counterpart to moment set 1:

$$\begin{aligned} Pr[d_{i'}|n_{i'}, a_k] &= \sum_j Pr[d_{i'}|n_{i'}, a_k, \nu_j] Pr(\nu_j|n_{i'}, a_k) \\ &= \sum_j Pr[d_{i'}|U_{i'} = 0, n_{i'}, a_k, \nu_j] Pr(U_{i'} = 0|n_{i'}, a_k, \nu_j) Pr(\nu_j|n_{i'}, a_k), \end{aligned} \tag{27}$$

where i' indexes simulated observations. Since our drilling game applies to owners of developed plots, only $Pr[d_{i'}|U_{i'} = 0, n_{i'}, a_k, \nu_j]$ comes from simulating drilling decisions on the maps and taking averages as just described. In case $n_{i'} = 1$, $Pr(U_{i'} = 0|n_{i'}, a_k, \nu_j) = 1$, so that $Pr[d_{i'}|n_{i'} = 1, a_k]$ is just the weighted average of simulated drilling rates across unobserved types, where the weights $Pr(\nu_j|n_{i'} = 1, a_k)$ are computed from the simulated data using Bayes' rule. In case $n_{i'} = 0$, we have

$$
\begin{aligned}
Pr(U_{i'} = 0|n_{i'} = 0, a_k, \nu_j) &= \int Pr(U_{i'} = 0|n_{i'} = 0, a_k, \nu_j, w_{i'})Pr(w_{i'}|n_{i'} = 0, a_k, \nu_j)dw \\
&\simeq \sum_{i'} \frac{\mathbb{1}(n_{i'} = 0, a_{i'} = a_k, \nu_{i'} = \nu_j)}{\sum_{i'} \mathbb{1}(n_{i'} = 0, a_{i'} = a_k, \nu_{i'} = \nu_j)}[1 + e^{-(\lambda_0+\lambda_1 w_{i'})}]^{-1}
\end{aligned}
$$

where the second line uses equation (25) and the fact that $w_{i'}$, as in Step 1 of estimation algorithm, is drawn from the conditional distribution $Pr(w_i|a_k, i \in \text{sample})$. We then apply equation (27) with simulated weights $Pr(\nu_j|n_{i'} = 0, a_k)$. The other moment conditions are constructed analogously.

We minimize the SMM criterion function, assembled from these moment conditions and the weighting matrix discussed above, using a downhill simplex method.

5.5 Structural estimation results

Table 4 reports the model parameter estimates along with their bootstrapped standard errors based on R replications of our estimation procedure. We can reject a Cobb-Douglas production function, which is nested within the CES, i.e., when $\delta = 0$. We also find, as expected, that $\lambda_1 > 0$, indicating that plots owned by wealthier farmers are more likely to be developed for groundwater. Overall, our model

Table 4: Parameter Estimates

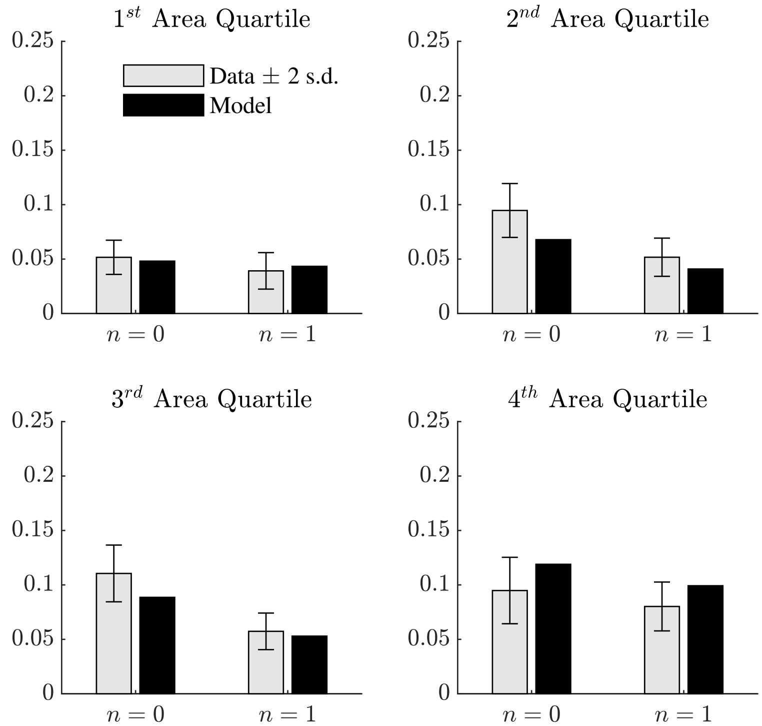
θ	α	δ	σ	λ_0	λ_1	s
18.87	0.18	0.73	1.29	-22.34	1.94	2.52
(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: Standard errors in parentheses. See equation (13) for definition of production function parameters (θ, α, δ); σ is standard deviation of random drilling shock; λ_0 and λ_1 are parameters of the probability of an undeveloped plot (see equation (25)); s is the shrinkage parameter to account for nonrandom sampling of reference plots on the basis of owner wealth.

matches the targeted moments reasonably well (see Figures 5-8). The model also does quite well in matching certain untargeted moments, as shown in Figures 6 and 8, although it tends to over-predict the fraction of plots with 2 borewells as well as the fraction with at least 4 neighboring borewells.

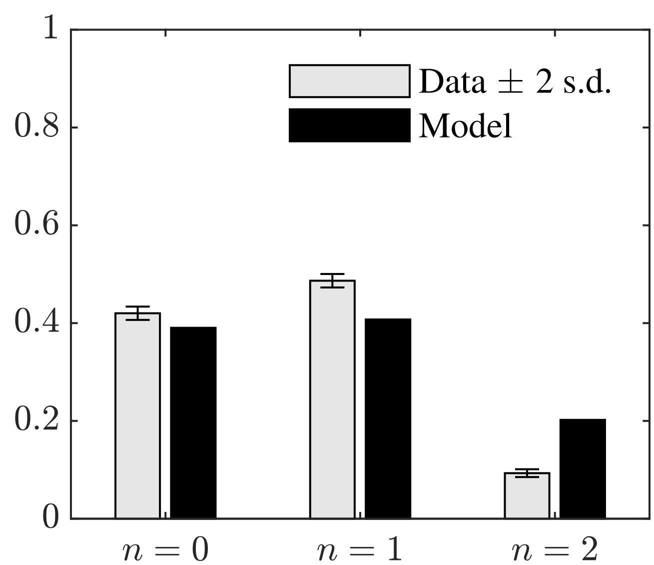
We next subject the structural model to a more exacting test of fit. Starting at a steady state on each village map, we simulate 10 five-year panels consisting of triplets $\{d_{it}, n_{it-1}, \mathcal{N}_{it-1} : t = 1, ..., 5\}$

Figure 5: Drilling rate by area quartile and number of functioning borewells



Note: All moments are targeted.

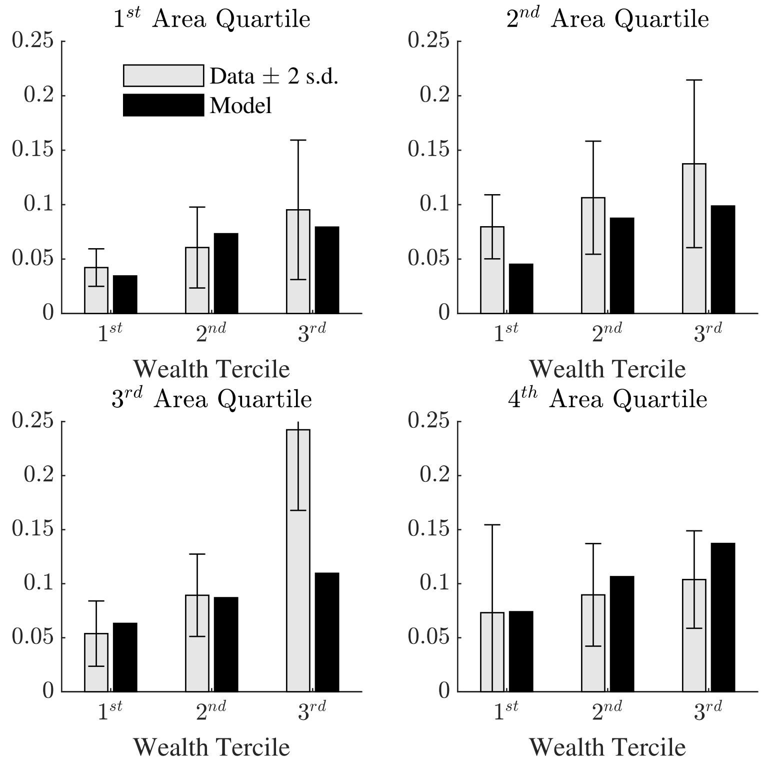
Figure 6: Fraction of plots with n functioning borewells



Note: $n = 0$ targeted; $n = 1, 2$ untargeted

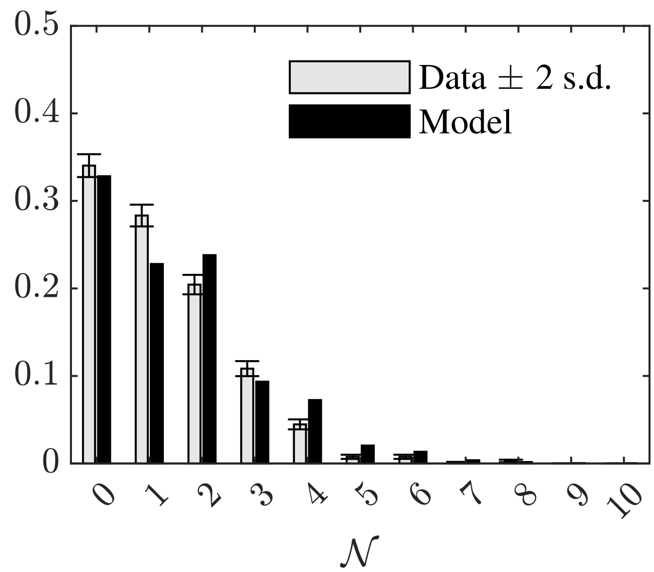
for every plot on the map that is assigned developed status (see Step 2 in Subsection 5.3), yielding 28,508 synthetic panels in total. We then randomly draw from this sample, by village, a number of

Figure 7: Drilling rate by area quartile and wealth tercile



Note: All moments are targeted.

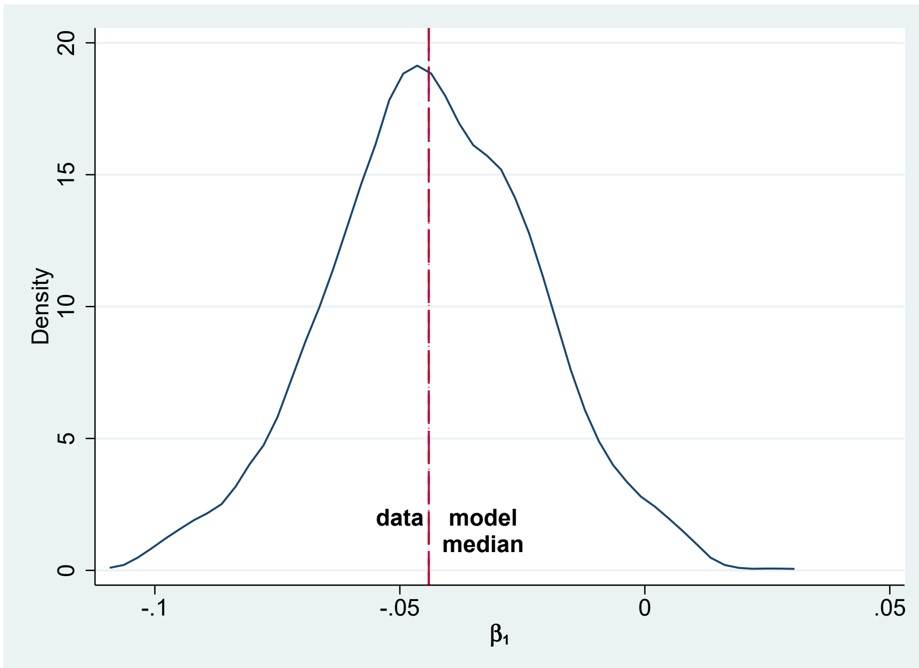
Figure 8: Fraction of plots with $\mathcal{N}$ surrounding functioning borewells



Note: $\mathcal{N} = 0$ targeted; $\mathcal{N} = 1, 2 \ldots 10$ untargeted.

"reference plots" equal to the number of reference plots contributed by that village to the actual sample of 1,057. Using 1000 draws of these simulated panels, we estimate a linear probability model version of the drilling reduced form reported (based on actual data) in column 5 of Appendix Table E.1.[39]

Figure 9: Test of model fit



Notes: Distribution of estimated coefficients from linear fixed-effects regressions of a drilling indicator on the number of functioning wells outside the reference plot using 1000 panels simulated from the structural model.

Since we have not exploited the correlation between d_{it} and $\mathcal{N}_{it-1}$ in estimating the structural model, we shall focus on β_1, the coefficient on $\mathcal{N}_{it-1}$ in the drilling regression (See equation (8)). In other words, β_1 is an untargeted moment, yet one most directly tied to the strategic behavior that our model is designed to capture. Figure 9 shows the distribution of β_1 estimates from the synthetic panels. Remarkably, the estimate of 0.0441 from the actual data virtually coincides with the median estimate from the 1000 replication samples. We view this as powerful corroboration of our model's validity.

[39]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, because misclassification error is not an issue in the simulated data, we do not use IV. Finally, since our model assumes (for tractability) that no drilling occurs on plots with two borewells, we drop observations with $n = 2$ in the estimation samples for both actual and synthetic data.

6 Counterfactuals

We now turn to the quantitative evaluation of counterfactual policies designed to address both the distortion induced by electricity subsidies and the negative well interference externality.

6.1 Preliminaries

With model parameter estimates in hand, we assess the implications for social welfare and groundwater development of an annual tax τ on each functioning borewell. Practically, τ could be implemented as a flat charge for maintaining an agricultural electrical connection, which, if set equal to the annual cost of electricity to run the pump (given a binding power supply rationing constraint), effectively counteracts the electricity subsidy; any excess of connection charge over electricity costs acts like a Pigouvian tax on borewells. In terms of the decision to drill, the annual *net* Rupee value of output under a counterfactual $\tau > 0$ is simply $\mathbb{E}[y(N,n)|R] - \tau n$. Once a borewell fails, we can think of its electricity connection as being disabled so that no further taxes are incurred.

As a prelude to the policy analysis, we use our structural model to simulate the value of existing and potential future groundwater development, or "welfare", in steady state on each map, and average the result across maps. We calculate private welfare per acre of (model-predicted) developed land as the expected discounted present value of agricultural output minus drilling costs. To be clear, our private welfare calculation takes account of well interference externalities inasmuch as it averages output across hundreds of adjacent plots in each simulated map. Using this approach, the incremental capitalized private value of groundwater development is 45,600 Rs/acre.[40] While this calculation does not net out the (sunk) cost of drilling extant borewells, it does implicitly discount the future income flows from these borewells by their failure probability and incorporates the option value of expected future drilling.

The social value of groundwater development is the private value, just discussed, minus the cost of electricity, which, though free to farmers, is not free to society. We find that this social value is only 14,500 Rs/acre. In other words, more than two-thirds of the private value of groundwater development is accounted for by the capitalized value of the electricity subsidy. Next, we recompute the village map equilibria under the counterfactual that 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 borewell density in steady state increases by more than 20 percent and that the social value of groundwater development rises to 22,100 Rs/acre. Thus, the negative externality diminishes the value of groundwater to society by around a third, a substantial economic burden.

[40]By way of comparison, median plot value in our study setting is 300,000 Rs per acre. Specifically, we collected information from farmers on the per acre market value of 2885 owned plots (including all reference plots for the adjacency survey) in 2017. The median plot does not have a functioning borewell.

6.2 Optimal tax on borewells

We now consider alternative borewell taxes τ, searching for the optimal tax τ^* that maximizes social welfare. Since well-drilling entails a negative externality, τ^* should exceed the annual cost of (currently freely provided) electricity, the difference being the Pigouvian component of the tax.

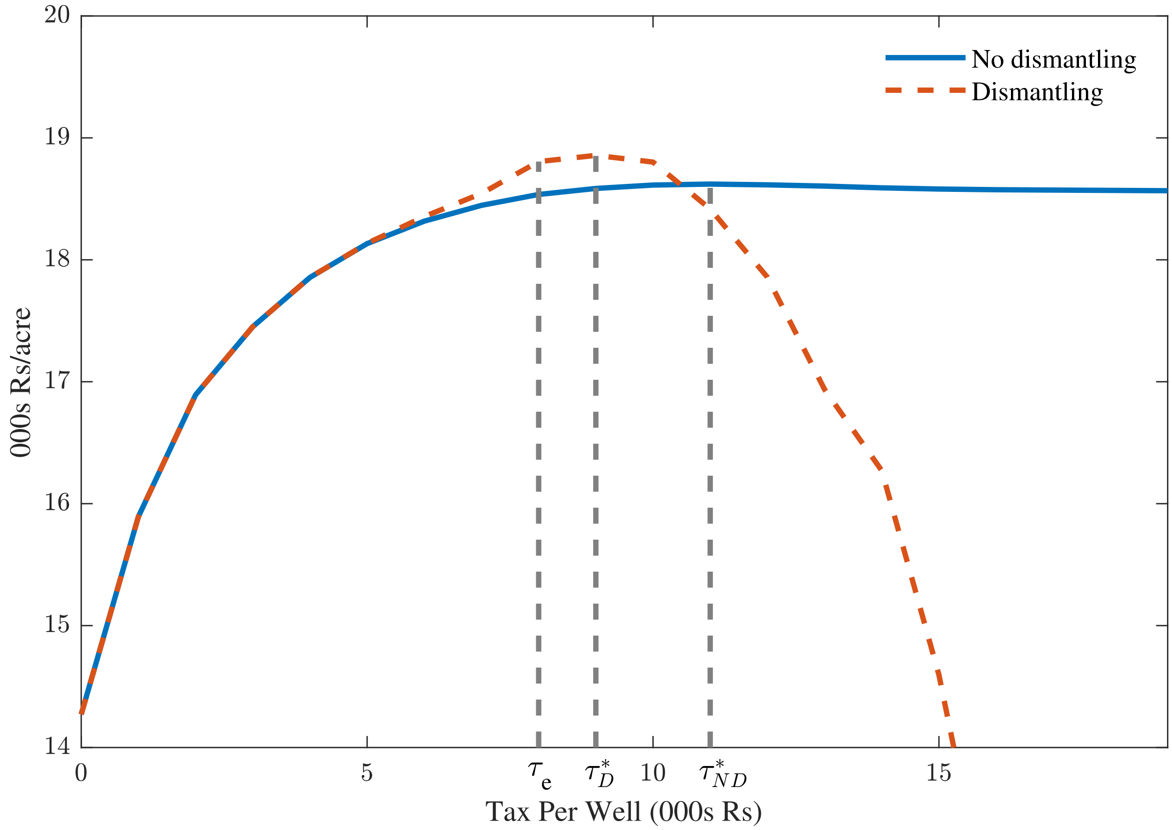
For each counterfactual borewell tax τ, we follow Domeij and Heathcote (2004) and compute social welfare along the transition path from the zero tax baseline. This calculation takes into account the "short-run", over which the existing stock of borewells is relevant and, therefore, allows us to compare alternative policy treatments of wells that have already been sunk.[41] We thus consider two regimes, the first in which all borewells (extant and newly sunk) are subject to taxation but extant wells can be dismantled at zero cost, and a second in which extant wells are "grandfathered" so that the tax only applies to new drilling (i.e., no dismantling of old borewells).

As shown in Figure 10, under either tax treatment of extant borewells, setting $\tau = \tau_e$, the annual electricity cost, would increase the social value of groundwater development to around 19,000 Rs/acre, or by 28%. In other words, the social cost of free electricity amounts to 4,500 Rs (65 US$) per acre, or about 1.5% of farmland value in our setting. The welfare gain is only slightly larger when the government charges all borewell owners for the electricity they use rather than charging only new borewell owners. To understand the magnitude of this deadweight loss, first note that removing the electricity subsidy is predicted to induce 0.16 fewer borewells per acre in the new steady state. Therefore, the deadweight loss is slightly more than 28,000 (= 4500/0.16) Rs per 'surplus' borewell. Meanwhile, the cost of electricity in expected present value terms is 85,000 Rs per borewell (i.e., 8500/0.1, where the numerator is the annual cost and the denominator is the discount rate), which is an upper bound on the deadweight loss per surplus borewell. In other words, 85,000 Rs is the marginal deadweight loss (from the *last* borewell sunk) under the subsidy policy, whereas 28,000 Rs is the average deadweight loss.

Turning to the first-best, when both new and extant borewells are subject to taxation, the social welfare maximizing tax, τ_D^* in Figure 10, equals annual electricity costs plus a Pigouvian premium of around 12% (above τ_e) to correct for the externality. When existing borewells are grandfathered, however, the social welfare maximizing tax equals τ_e plus a roughly 30% Pigouvian premium (about 2,500 Rs per borewell per year); the higher premium is required because the marginal externality cost of a new borewell is higher when extant borewells still operate. Nevertheless, both tax policies curb excessive drilling and thus achieve very similar social welfare along their respective transition paths. Grandfathering of existing borewells is, however, the more politically palatable policy as no farmer

[41] By contrast, a comparison of steady states with and without a borewell tax provides a long-run perspective, tantamount to comparing alternative histories of groundwater development starting from an initially clean slate. The short run analysis that we pursue here is arguably the more salient for policy-making purposes.

Figure 10: Social welfare under alternative borewell taxes

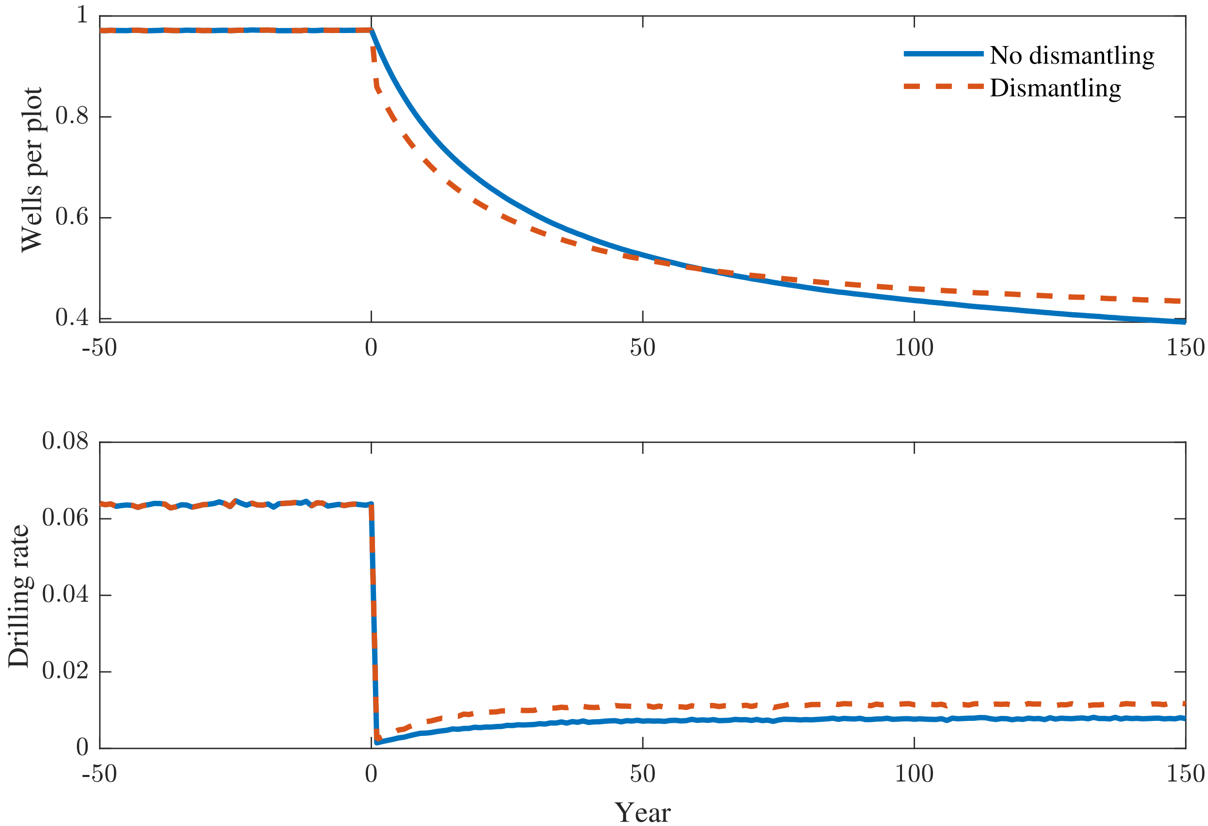


Notes: Each point on the solid (dashed) curve represents the social welfare along the transition path from the benchmark zero-tax economy to the steady state under the corresponding (on the x-axis) counterfactual tax on newly drilled (both new and old) borewells: $\tau_e = 8.5$ is the tax that recovers electricity costs; $\tau^*_D = 9.5$ is the optimal tax applied to all borewells (dismantling); $\tau^*_{ND} = 11.0$ is the optimal tax applied only to newly drilled borewells (no dismantling).

would suffer a capital loss on sunk investments.

Lastly, Figure 11 shows the transitional dynamics for the optimal tax under each treatment of extant borewells. When all borewells, old and new, are taxed, a large fraction of existing wells are dimantled when the policy is first implemented, hence the vertical drop in wells at time 0 (dashed curve, top panel). Since those borewells that continue to be operated tend to be on high flow type plots, they fail infrequently, so that the subsequent decline in borewell numbers is quite slow. By contrast, when only new borewells are subject to taxation, the initial decline in well density occurs more slowly as it comes entirely through stochastic failure. Eventually, however, the number of wells per plot under the optimal tax targeted to new drilling falls below that of universal borewell taxation. This is because, compared to the former scenario, the optimal tax is considerably higher, implying a lower steady state drilling rate (bottom panel of Figure 11).

Figure 11: Transitional dynamics



Notes: The solid curve in the top (bottom) panel shows the number of borewells per plot (drilling attempts per plot) along the transition path from the benchmark zero-tax economy to the steady state under τ_D^*, whereas the dashed curves do the same for the transition to the steady state under τ_{ND}^*.

7 Conclusion

Our main goal in this paper has been to assess the social cost of a current government policy of providing free electricity to groundwater extractors in south India. To tackle the highly localized (and economically important) well interference externality, we have developed a tractable strategic equilibrium model of well-drilling on a large map of agricultural plots along with a novel estimation strategy, an approach potentially applicable to dynamic discrete network games across a wide range of settings. Counterfactual analysis based on the model reveals a social cost of free electricity amounting to 4,500 Rs (65 US$) per acre in present value terms, or around 1.5% of farmland value.

The presence of externalities also raises the broader question of the optimal tax on borewells, which would not only eliminate the distortion from the electricity subsidy but would also maximize social welfare. In taking into account how such a tax affects existing borewell investments in the transition to the new steady state, our counterfactual analysis is uniquely relevant for policy. Indeed, we find that only a modest Pigouvian tax would be needed to correct for the externality when both new and existing borewells are subject to taxation, whereas the requisite Pigouvian tax increases by a factor of about 2.5 when it applies only to new borewells. That said, the latter policy (in practice, charging only for new electrical connections) makes the most sense on political-economy grounds as, with minimal loss in social welfare relative to the broader tax, it avoids a capital levy on existing well owners, an influential interest group in rural India.

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.

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.

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.

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.

Rosenzweig, M. R. and K. I. Wolpin (1993): "Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India," *Journal of political economy*, 101, 223–244.

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.

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

A 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.

B Descriptive figures and tables

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

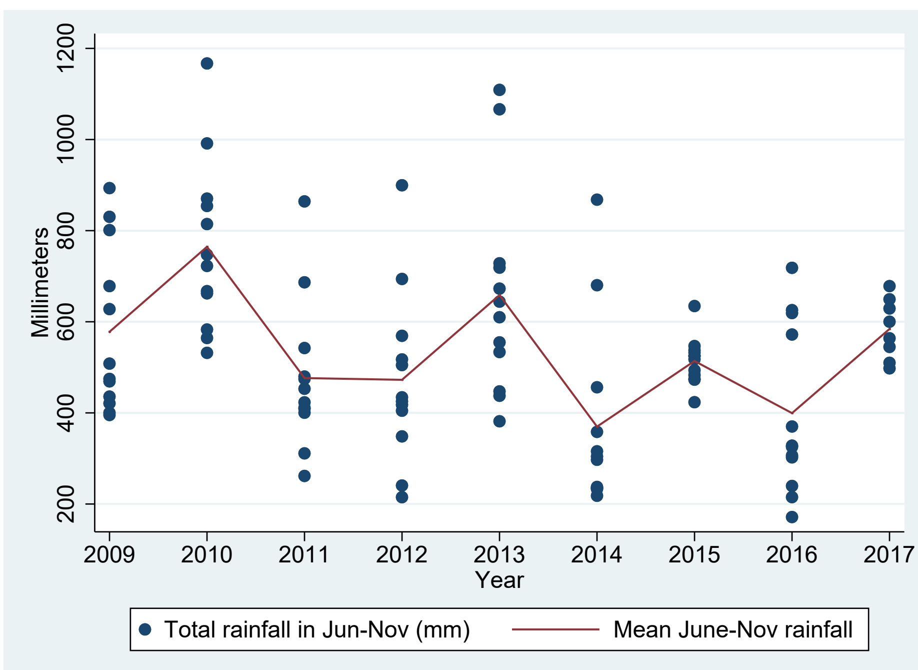


Table B.1: Well Failure by Year

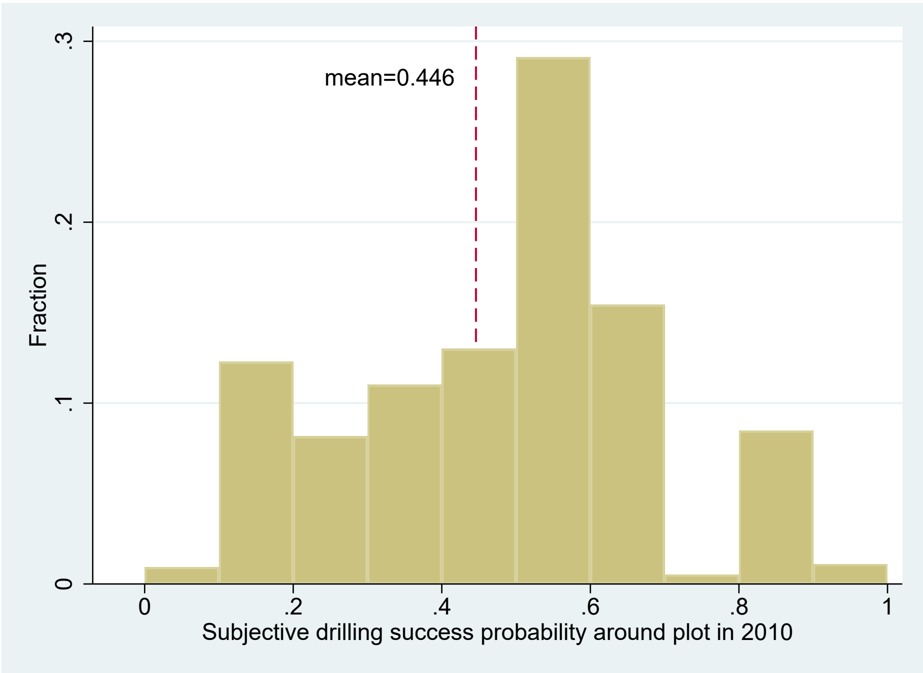
Year	Functional	Failed	Total
2012	559	47	606
	(92.2)	(7.8)	(100)
2013	556	41	597
	(93.1)	(6.9)	(100)
2014	527	53	580
	(90.9)	(9.1)	(100)
2015	512	33	545
	(93.9)	(6.1)	(100)
2016	489	34	523
	(93.5)	(6.5)	(100)
Total	2,643	208	2,851
	(92.7)	(7.3)	(100)

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

Table B.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

Figure B.2: Drilling success probability



C 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 C.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 C.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).

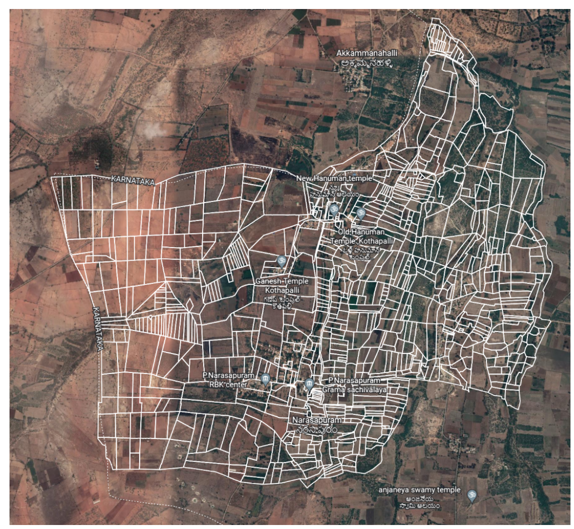
D Cadastral Maps

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 digitize maps to identify the adjacency of each plot that will be used in the structural estimation.

Figure D.1: Village Muddiredipalle

((a)) Original Cadastral Map



((b)) Digitized Cadastral Map

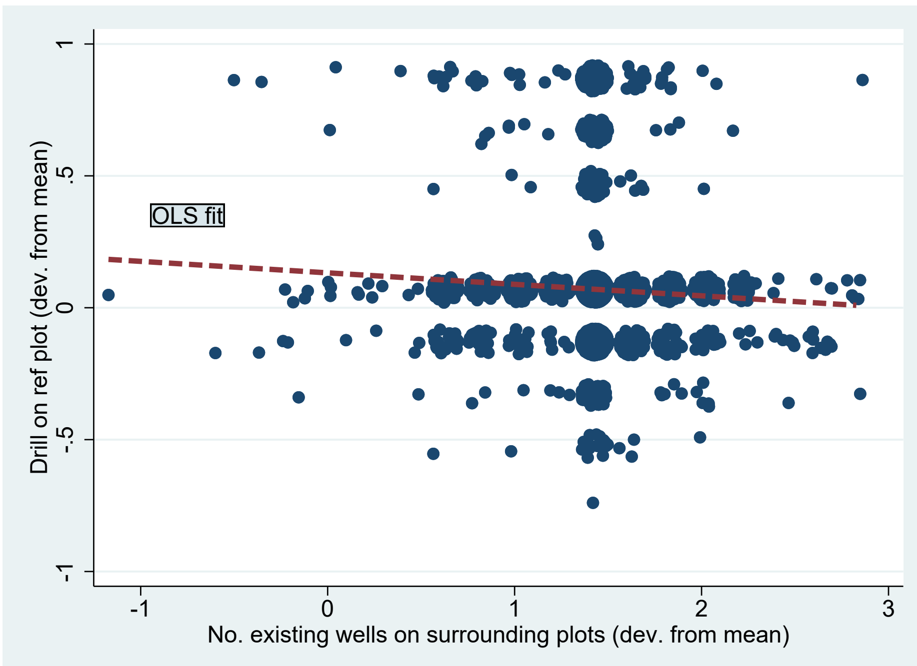
E Drilling propensity: Linear FE-IV estimates

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

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

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), $\mathcal{N}^E_{it-1}$, as an instrument. Figure E.1 shows the within reference plot (i.e., fixed effects) regression of d_{it} on $\mathcal{N}^E_{it-1}$, which is essentially the reduced form corresponding to our IV regression of equation E.1.

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



Column (1) of Table E.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 $\mathcal{N}_{it-1}$ on the instrument $\mathcal{N}^E_{it-1}$ 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 $\mathcal{N}^E_{it-1}$ 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.[42] 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 E.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 ($\mathcal{N}$)	-0.0150		-0.0482**	-0.0485**	-0.0441
	(0.0108)		(0.0213)	(0.0199)	(0.0198)
Good Monsoon (R)	-0.00340	0.000760	-0.00248	-0.00257	
	(0.00881)	(0.00775)	(0.00884)	(0.00866)	
No. exist wells exc ref plot		0.900***			
		(0.0285)			
1 func well on ref plot				-0.243***	
				(0.0241)	
2 func wells on ref plot				-0.447***	
				(0.0538)	
Ref plot FE	YES	YES	YES	YES	YES
Observations	5,285	5,285	5,285	5,285	4,837
R^2	0.226	0.969	-0.002	0.067	-0.0024

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.

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