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

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

February 14, 2024

Abstract

Borewells for groundwater extraction have proliferated across South Asia, encouraged by subsidized electricity for pumping. Because borewells operating near one another experience mutually attenuated discharges and higher failure rates, farmers deciding whether and when to drill interact strategically with potentially many neighbors, both directly and indirectly through the spatial network of agricultural plots. We develop a tractable approach to estimating a structural model of well-drilling as a dynamic discrete network game and implement it using plot-level panel data from two states of southern India. Using the estimated model, we compare the current regime of free (but rationed) electricity against an annual tax on all functioning borewells that fully defrays electricity costs. We find that the counterfactual policy, by reining in over-drilling, eliminates a deadweight loss of 170 US$ per acre of land with groundwater potential, 30% of the fiscal cost of the subsidy. Further, we find that taxing borewells at a rate 18% higher than annual electricity costs (to address the negative externalities) is socially optimal. Our estimates also suggest a practical compensation scheme to mitigate potential farmer opposition to electricity cost recovery.

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 cadastral 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. Pedro Mira acknowledges funding from the Spanish Ministry of Science and Innovation (grant PGC2018- 097598-B-I00).

1 Introduction

Land-specific investments often occur within spatial networks characterized by local spillovers. Developing a given parcel may thus substantially affect the returns to developing closely neighboring land even while having negligible direct spillovers further away. Such local externalities, moreover, induce strategic interactions, with each landowner's irreversible investment depending on how their neighbors have invested in the past or might invest in the future. Each of these neighbors, in turn, is influenced by the decisions of their neighbors, and so on. Given this network-game structure, quantifying how a policy change would affect aggregate welfare in general equilibrium presents a significant methodological challenge. This paper develops a tractable approach to estimating policy counterfactuals for the case of borewell investment in southern India, a context where both local spillovers and government intervention in private investment decisions are particularly salient.

Groundwater has become the dominant source of irrigation in India, driving increased agricultural intensification (Jain et al., 2021) and rising rural income (Sekhri 2014). To extract this resource, millions of borewells have sprung up in recent decades, most equipped with submersible electric pumps (Shah 2010; Jacoby 2017). As a borewell operates, the water table around it drops, creating a conical drawdown region centered on the pump. If two borewells are close enough to one other, their respective cones of depression overlap, significantly reducing the water flow from each. An additional functioning borewell nearby thus lowers the discharge of any given borewell, and may ultimately lead to its failure, i.e., when discharge is too low to warrant any cultivation whatsoever.[1] In the dry season, during which groundwater is the sole source of irrigation in most of India, well interference becomes an important, albeit highly localized, common property externality.

Most Indian states provide farmers free or highly subsidized electricity to run their pumps, artificially inflating the economic returns to well-drilling.[2] Groundwater development has thus devolved into "drilling for subsidies", a form of rent-seeking in which smallholders sink costly wells that would not be economically viable absent these policy-induced distortions (Badiani-Magnusson and Jessoe, 2018). By driving the gross return for the marginal borewell below the private cost of drilling plus the fiscal cost of the subsidy, government policy creates welfare losses on top of those induced by any common property externalities. Were these groundwater externalities not highly localized, it would be reasonable to assess the social cost of over-drilling using a partial equilibrium framework. In the presence of well interference, however, only a general equilibrium framework that captures both the costly externality and the

[1]Blakeslee et al. (2020) find large negative economic impacts of well failure in the South Indian state of Karnataka.

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

resulting strategic interactions between neighboring farmers can generate valid policy counterfactuals.

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

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

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

Based on these results, we calculate the social value of land suitable for groundwater development as the expected discounted present value of its agricultural output net of drilling costs and the cost (to society) of the electricity to run the pumps of any functioning borewells. We find that this social value is practically zero; in other words, the private value to landowners consists entirely of the capitalized value of the electricity subsidy. We also find that charging all borewell owners fully for electricity through an annual flat tax would increase social value by 11,000 Rs/acre (US$ 170), an implied social cost (deadweight loss) amounting to 30% of the fiscal cost of the subsidy. We further consider the optimal Pigouvian borewell tax that would eliminate deadweight losses from free electricity provision and from the well interference externality. While we obtain an optimal tax 18% higher than the cost of electricity, social welfare is almost the same as under a policy of electricity cost recovery. Lastly, we examine the distributional implications of electricity cost recovery, and borewell taxes more generally, suggesting a simple compensation scheme to reduce inequities and build political support for reform.

Contribution Despite the seeming ubiquity of local investment spillovers, especially in natural resource economics, the literature on dynamic discrete network games is extremely sparse. Indeed, the only other attempt to estimate such a model to our knowledge is that of Hodgson (2021) in the context of oil field exploration, where the externality is informational and the key inefficiency is thus one of free-riding rather than of over-investment. The remaining empirical applications of network games are exclusively static (see, e.g., Acemoglu et al. 2015, Xu 2018, König et al. 2017). A static model in which all plot owners drill borewells simultaneously would not account for well failure, an important empirical phenomenon in our setting that is inherently dynamic. Moreover, even in the absence of well failure, only a dynamic model allows us to exploit data on drilling choices conditional on the current number of functioning borewells – i.e., the dynamic decision rule – to identify the own productivity effect of neighboring borewells (the static externality). Lastly, a one-period model could not speak to the short-run distributional, and hence political-economy, implications of a borewell tax. A dynamic model, by contrast, allows us to study the differential impacts of such a policy change on landowners with and without functioning borewells at baseline, thereby informing the design of practical compensation schemes.

As the first paper to empirically incorporate the well interference externality into a full dynamic general equilibrium model of drilling decisions, this paper also contributes to the economics of groundwater extraction, a literature focused almost exclusively on the United States and, to a lesser extent, on the world's largest groundwater user, India. Pfeiffer and Lin (2012) addresses well interference externalities in the High Plains Aquifer of Kansas whereas Sears et al. (2022) studies strategic interactions among neighboring extractors in California. Both of these papers are concerned with the intensive margin (pumping) as opposed to the extensive margin (drilling) and neither invokes a spatial network game. Groundwater research set in India focuses on the the deep alluvial aquifers of the northwest, where 'mining' of fossil groundwater is a serious concern (see Fishman et al. 2011). In this context, the calibration study of Sayre and Taraz (2019) considers both groundwater pumping and investment in deeper wells within a partial equilibrium framework, i.e., ignoring well interference externalities. Ryan and Sudarshan (2022) econometrically estimates the welfare cost associated with over-pumping in Rajasthan. In focusing on the *non-local*, aquifer-wide, pumping externality as well as in taking the number of borewells as fixed, their paper abstracts from both well interference and from drilling costs. Ryan and Sudarshan (2022) finds that electricity rationing to agriculture leads to roughly the socially optimal quantity of groundwater pumping on average. Similar rationing in our study areas of AP and

Telangana undoubtedly also limits over-pumping, rendering this intensive margin distortion small in comparison to the extensive margin distortion that we emphasize.

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

2 Setting and Data

2.1 Context

Before its partition, Andhra Pradesh was one of the most important agricultural states of India, accounting for about 7 percent of gross cropped area nationally, with groundwater supplying roughly half of its irrigation. As argued by Kumar et al. (2011), however, the economic efficiency of groundwater extraction in AP has been falling substantially, with the tripling in the number of borewells to more than 1.5 million from 1995-2010 (see Jacoby 2017), leading to high rates of well failure, lower area irrigated per well, and higher energy requirements for groundwater pumping due to well interference. Power supply to agriculture for running electrical pumps has, meanwhile, become a political issue all over India. In 2004, a newly elected government in AP abolished flat rate electricity charges to farmers, which had previously covered just 11 percent of the cost of provision, and introduced free agricultural power, a move swiftly followed by the major states of Tamil Nadu, Karnataka, and Punjab.[3] Farmers in AP and Telangana typically run their pumps continuously during the fixed number of hours (7-9) per day when this free electricity is made available for agricultural use (Fishman et al., 2023).

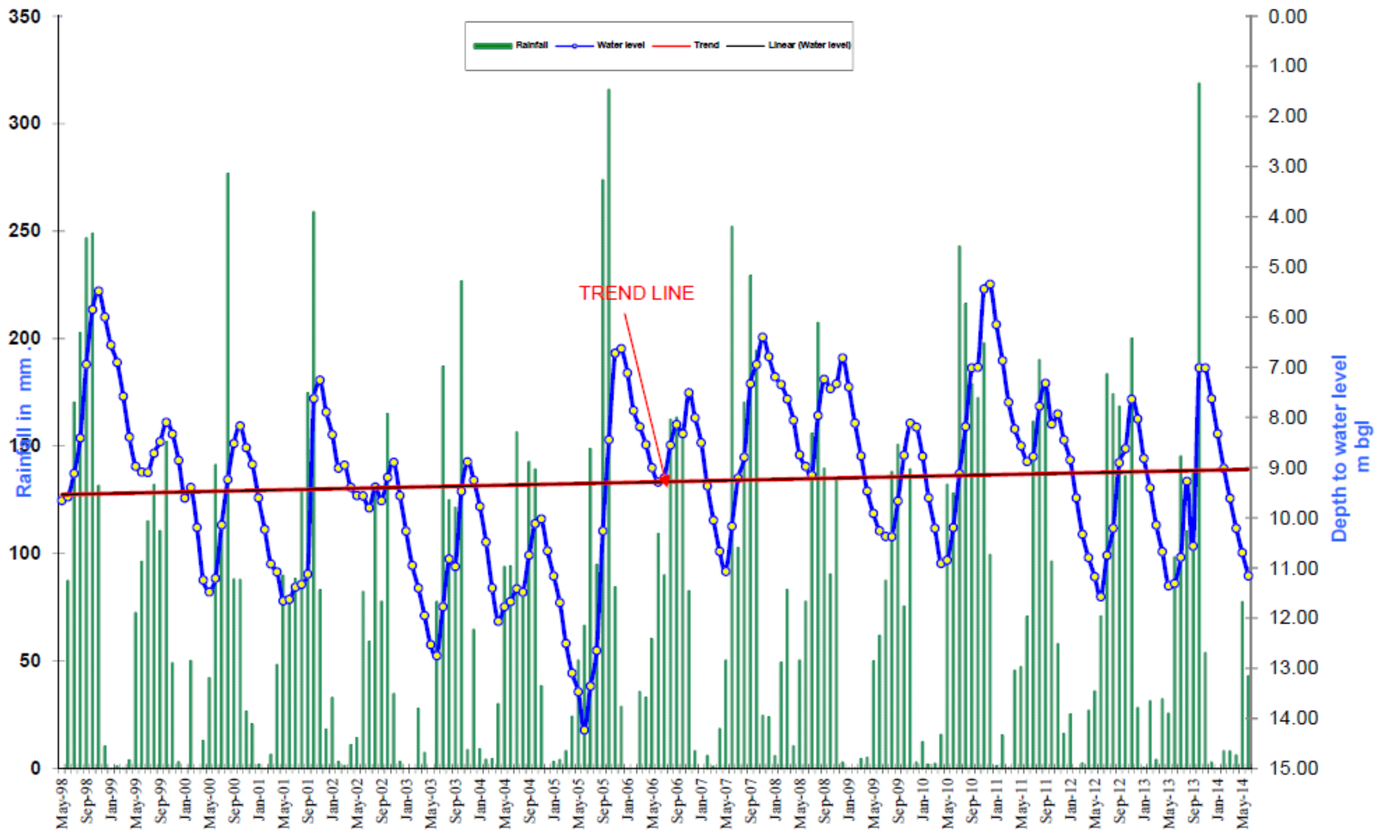
Much of South India is underlain by shallow hard rock aquifers with limited groundwater storage capacity. Recharge from monsoon rains is thus largely depleted through pumping during the subsequent dry season; in contrast to northwest India, there are no deep groundwater reserves available to mine. As 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.[4] Because of their low

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

[4]Hora et al. (2019) argues that such water table trends are biased upward by relying on surviving (i.e., non-failed)

transmissivity (velocity of horizontal groundwater flow), well interference is accentuated in hard-rock aquifers; in our setting, an interwell spacing of at least 250 meters is recommended to avoid interference effects (Chandrakanth 2015). Blakeslee et al. (2020) detail the process of well failure in hard-rock aquifers, highlighting *local* hydrogeological features, i.e., sub-surface fractures fed from different sources of recharge, as opposed to aquifer-wide depletion, a process hastened by competing extraction among neighbors. In sum, well interference is the predominant groundwater pumping externality in our setting, one that is both localized and static, affecting only current groundwater availability.

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).

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

observation wells to measure groundwater levels across time. Indeed, our analysis of well failure in Appendix E is consistent with a secular, but rather slow, decline in water tables in our study area.

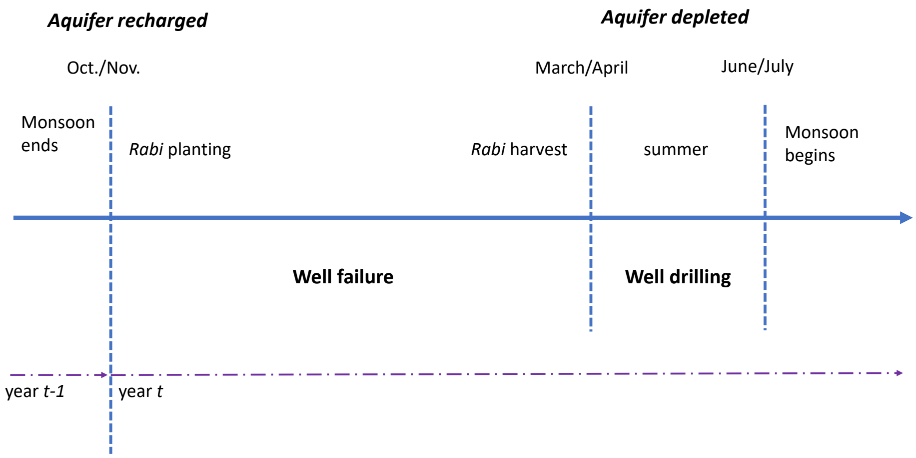
in these drought-prone districts is extremely limited compared to districts in the intermediate range of annual rainfall and, especially, to those in water-abundant coastal AP. Only farmers with access to a functioning borewell can cultivate during the dry (rabi) season, typically growing groundnut,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

2.2 Representative plot sample and retrospective panel

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

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

Figure 2: Timeline



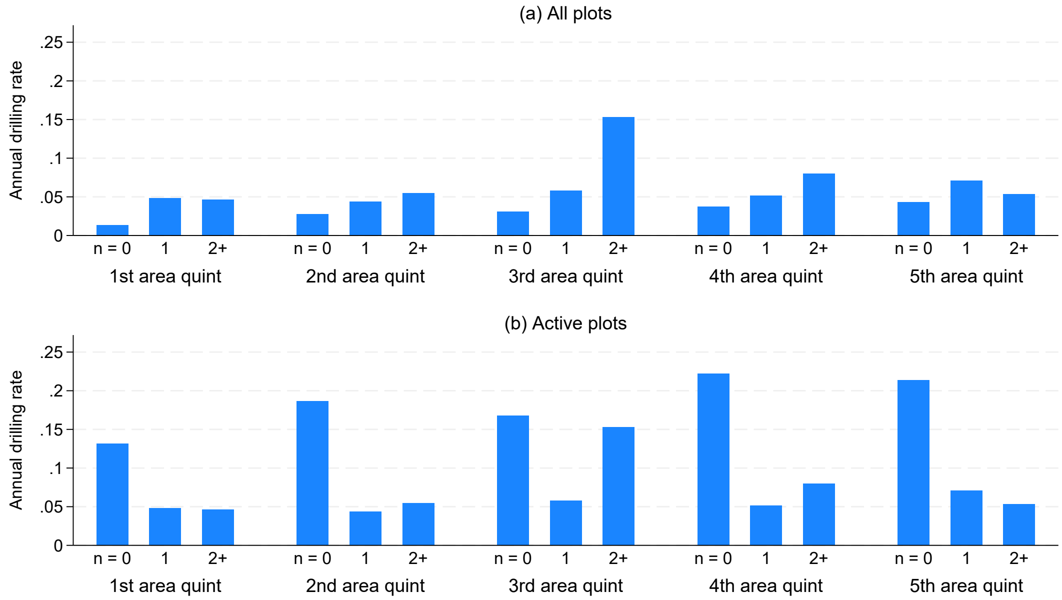
there is the potential for truncation of well drilling attempts and failures occurring later in 2017.[5]

Drilling patterns There were 526 drilling attempts made on 437 plots between 2012-16, only 197 (37.5%) of which were successful (i.e., resulted in a functional borewell). Panel (a) of Figure 3 shows annual drilling rates by the number of functioning borewells and plot area quintiles. While the propensity to drill is *higher* on plots with more functioning borewells, we conjecture that this reflects heterogeneity in groundwater potential; where drilling is profitable, there is both more drilling *and* more functioning borewells. To explore further, we restrict attention in panel (b) to those plots on which at least one drilling attempt was made during 2012-16 or which already had at least one functioning borewell in 2012. For these "active" plots (39% of the representative sample), the owners of which evidently believed had the potential for groundwater development, we observe a negative relationship between functioning borewells and drilling, consistent with diminishing marginal returns to investment.

Liquidity constraints Results reported in Appendix B show no link between the pre-sample wealth of the plot owner and their propensity to drill from 2012-16. This finding, along with other descriptive evidence adduced in Appendix B, suggests that it is reasonable to abstract from financial frictions in the borewell investment model of Section 4.

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

Figure 3: Drilling attempts



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

2.3 Adjacency survey and retrospective panel

Our analyses of the determinants of drilling (Subsection 3.3) and of borewell failure (Subsection 3.4) rely on an adjacency survey that was part of the 2017 instrument. An *adjacency* is defined as the set of all agricultural plots contiguous to the reference plot, inclusive of it. We assume that only other borewells operating in the adjacency create interference effects on the reference plot, imposing a negative externality. Put differently, borewells outside the adjacency do not influence the flow and failure of borewells in the reference plot. Given the typical size of plots and the range of well interference effects in our setting, we believe that this assumption is sensible.[6]

All 1,057 farmers with an eligible reference plot were administered an adjacency survey. A plot was eligible if at least one drilling attempt had been made in the last seven years either on the plot itself or

[6]In a chessboard configuration of identical plots averaging one hectare (as in our data) with borewells located at the center, the distance between a borewell in the reference plot and one elsewhere in the adjacency would be 100-140 meters, well within the range of interference effects mentioned by Chandrakanth (2015). 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.

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

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

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

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

2.4 Borewell failure panel

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

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

2.5 Borewell flow panel

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

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

3 Preliminary Estimation

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

3.1 Econometric issues

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

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

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

A key concern is that ν_i may be correlated with (N_{it}, R_{mt-1}). Since nonlinear probability models do not lend themselves to fixed effects approaches (except in some special cases), we employ *correlated random effects* (CRE) under the assumption that N_{it} and R_{mt-1} are strictly exogenous conditional on ν_i (see, e.g., Wooldridge 2010).[7] In particular, let

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

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

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

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

Selection bias Our estimation sample for the flow/failure model is selected on the basis of whether at least one borewell was successfully sunk on the plot prior to or possibly, in the case of failure, during the panel period. Since the pre-existence of a borewell on a plot could be correlated with the unobserved heterogeneity driving borewell flow and/or failure on that plot, selection bias is a concern. However, insofar as ε_{it} is iid over time, and given that entry into the failure panel begins at least a "year" after the borewell is sunk (see Figure 2), this sample selection is strictly exogeneous; in other words, conditional on ν_i, selection is uncorrelated with idiosyncratic flow/failure shocks over the panel period. In these circumstances, the nonlinear CRE estimator is robust to selection bias (Wooldridge 2019).

Misclassification error The reported number of functioning wells on *neighboring* plots, $\mathcal{N}_{it}$, may be subject to recall error, leading to a non-classical form of measurement error.[8] We thus assume, quite plausibly, that the number of *existing* neighboring wells $\mathcal{N}_{it}^E$ is accurately observed. Ignoring n_{it}, which is accurately measured by assumption, we thus want to estimate the probability that some outcome Y_{it}

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

[8] 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.

depends upon the true number of functioning wells in the adjacency $Pr(Y_{it}|\mathcal{N}_{it})$. Although $\mathcal{N}_{it}$ is not perfectly observed, we know that

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

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

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

where p is the underlying annual probability of well failure.[9] The misclassification error model (MEM) estimator then assumes that the likelihood contribution conditional on unobservable μ is

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

3.2 Determinants of drilling

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

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

In addition to allowing for separate effects of own (β_2 and β_3) and neighboring (β_1) borewells, equation (8) includes the vector of controls Z_i consisting of log reference plot area, number of plots in the adjacency, and mandal dummies. Well interference induces strategic substitutability between drilling decisions of neighboring plot owners, implying that $\beta_1 < 0$.

[9] More precisely, p should be thought of the average failure rate of borewells in the adjacency of reference plot i, excluding those on the reference plot itself. We show empirically below that p is a function of the number of functioning borewells in the relevant neighborhood. In the context of equation (5), the relevant neighborhood is that around each of the plots *adjacent* to the reference plot, about which we have no adjacency-level data. Given this, we approximate p using the district average of the actual failure rate, yielding $\hat{p} = 0.104$ for Anantapur and $\hat{p} = 0.052$ for Mahabubnagar.

Identification While the strategic substitutability parameter β_1 resembles a peer effect, causal identification is not as challenging as implied by Manski (1993). In particular, since $\mathcal{N}_{it}$ reflects past (as opposed to current year) drilling by neighbors, and we are using panel data methods, 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.[10] While identification does break down if ε_{it} is both spatially *and* serially correlated, 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 the likelier scenario of positive spatial and serial correlation of drilling shocks, our estimate of β_1 would be biased upward, i.e., toward zero, and thus *away* from strategic substitutability.

Results Table 1 reports the determinants of annual drilling decisions on the 1,057 reference plots covered by the adjacency survey. In column 1, a CRE logit specification that ignores misclassification error shows a negative but insignificant effect of $\mathcal{N}_{it}$ on drilling (first row). After correcting for misclassification using the MEM estimator, we find (column 2) that having more functioning wells in the adjacency outside of the reference plot *does* significantly reduce the likelihood of further drilling.[11] This is strong evidence of strategic substitutability between neighbors' drilling decisions.

Table 1 also shows that having more functional borewells on the reference plot substantially reduces well-drilling by the plot owner, presumably due to diminishing returns to groundwater in production and to potential interference among own borewells; the predicted annual drilling rate fall from 0.132 to 0.044 to 0.017 in going from, respectively, 0 to 1 to 2 reference plot borewells. A good past monsoon does not significantly affect the propensity to drill, as would be expected if rainfall were iid across years and thus not predictive of future rainfall. Drilling is more likely on larger plots, but does not depend on how many plots there are in the adjacency, which supports a simplified state-space $(\mathcal{N}, n)$ for the reference plot owner's investment decision (see Section 4.2 below). The inclusion of dummies for the 12 mandals (col. 3) has little effect on any of the results in Table 1.

For each specification, we perform a likelihood ratio test against the null of no unobserved heterogeneity at the reference plot level, referring specifically to the distribution of unconditional heterogeneity ν_i (see the next subsection for details of the testing procedure). We strongly reject the null of no heterogeneity (bottom row of Table 1).[12] This finding suggests that unobserved heterogeneity explains part

[10] Pfeiffer and Lin (2012) estimates the effect of a neighbors' groundwater pumping on own pumping behavior using an instrumental variables strategy. In this case, however, the strategic decisions are simultaneous and the econometric approach is essentially cross-sectional.

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

[12] The p-values account for testing on the boundary of the parameter space; i.e., they are one-half of the probability

Table 1: Determinants of Drilling

	(1)	(2)	(3)
No. func. wells outside ref. plot ($\mathcal{N}$)	-0.143	-0.549	-0.599
	(0.170)	(0.211)	(0.203)
1 func. well on ref. plot ($\mathbb{1}_{n=1}$)	-1.083	-1.150	-1.092
	(0.195)	(0.199)	(0.195)
2+ func. wells on ref. plot ($\mathbb{1}_{n=2}$)	-2.204	-2.350	-2.065
	(0.373)	(0.383)	(0.375)
Good monsoon (R)	-0.0622	-0.0494	-0.0608
	(0.171)	(0.172)	(0.186)
Log(ref plot area)	0.366	0.372	0.337
	(0.0850)	(0.0858)	(0.0883)
No. of plots in adjacency	-0.0332	-0.0381	0.0170
	(0.0631)	(0.0636)	(0.0629)
Mandal dummies	No	No	Yes
Estimation method	CRE	CRE-MEM	CRE-MEM
Log-likelihood	-1296	-1293	-1271
p-value (H_0 : No heterogeneity)	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 Section 3.1) and MEM to misclassification error model (Section 3.2).

of the cross-sectional correlation between drilling and the number of functioning wells in the adjacency and should thus be addressed in the structural estimation.

3.3 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.[13] 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.

that a chi-squared with 1 degree of freedom is greater than the LR test statistic.

[13]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.

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 $c_6 = \infty$.[14]

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

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

where $z_{it}^F(\xi)$ is a linear index for failure, F_{it} is a binary failure indicator, τ_i is the year that the borewell first enters the panel (or 2012, whichever comes last), T_i is the last year the borewell exists in the panel (or 2016, whichever comes first), and ξ is the unobserved heterogeneity in well failure. As also mentioned in Section 2.4, for reference plots with multiple wells, only the first one sunk is included in the failure panel. Allowing multiple borewells on a plot would lead to a violation of strict exogeneity due to correlation between N_{it} and the failure shock.

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

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

[14] Our measure of the number of neighboring borewells differs between flow and failure estimation datasets. In the former case, owners of functioning wells were asked about the number of functioning borewells within a 100 meters radius of the reference plot, which is not precisely the same as the number in the adjacency. In practice, however, the total number of borewells within 100 meters averages 2.40 as compared to 2.36 in the average agency (for the 588 reference plots with both measures available). Since the $\mathcal{N}_{it}$ used in the flow estimation is the contemporaneous (rather than retrospective) report of the respondent, we assume no misclassification error in (9).

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.[15]

Table 2: Joint Flow-Failure CRE Estimation

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

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

To estimate the probabilities of the five well flow states, $\pi_k(N, R, \nu^f)$, and the failure probability,

[15] 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).

$\pi_F(N, R, \nu^F)$, 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).

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 $H_0 : \sigma_\nu = \kappa_\nu = 0$.[16] 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.

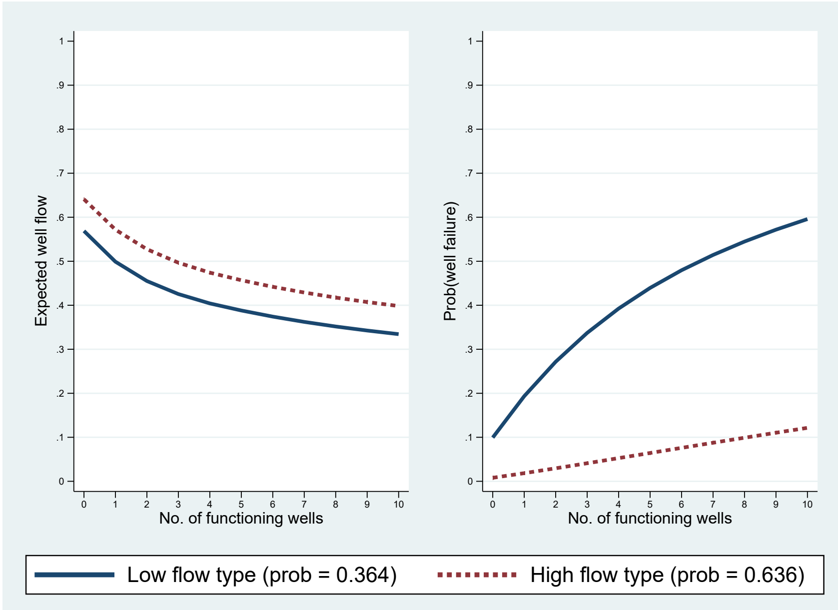
The top panel of Table 2 reports the coefficient estimates from Step 1. Column 1 ignores misclassification error, column 2 corrects for misclassification error using the MEM approach, and column 3 adds mandal dummies to the column 2 specification. Corroborating the well interference externality, we find that having more borewells in an adjacency depresses flow and makes failure of the reference well more likely. This latter effect, however, only emerges with the MEM estimation in columns 2 and 3. Also, having had a good previous monsoon improves well flow but does not have a significant effect on failure, consistent with our interpretation of well failure as an absorbing state, independent of the vagaries of the monsoon.[17] Including mandal dummies (in both flow and failure indices) shrinks the estimated scale of unobserved heterogeneity σ_ν from 1.39 to 0.31. Lastly, only the MEM specifications in columns 2 and 3 show the expected negative correlation between flow and failure heterogeneity and, in both specifications, we strongly reject the null of no unobserved heterogeneity (Step 3).

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

[16] 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.

[17] Once again, in Appendix F, we obtain qualitatively similar results using linear models for both flow and failure separately with plot fixed effects and instrumental variables to correct for measurement error (in the case of failure).

Figure 4: Expected flow and failure probability by type



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

types. By contrast, both the rate of well failure and the marginal effect of N on failure (extensive margin externality) are higher for the low flow type (probability = 0.364) than for the high flow type (probability = 0.636).

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

4 A Model of Borewell Investment

In this section, we present a dynamic equilibrium model of borewell investment on a large plot network. For reasons discussed in Section 2, we assume that plot owners do not face liquidity constraints when deciding each year whether or not to drill.

4.1 Preliminaries

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

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

where (θ, α, δ) are parameters, a_i is plot area, and q_{it} is well discharge or flow. Yearly flow is stochastic and thus unknown to the farmer prior to drilling and has a discrete distribution with K points of support $\{q_{it1}, ..., q_{itK}\}$ each with probability π_{itk}. Along with constant elasticity of substitution, production function (13) imposes constant returns to scale (CRS); i.e., output per acre depends only on flow per acre.[18] The scale parameter θ converts physical output into 2017 Indian rupees (Rs).

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

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

The well interference externality, estimated in Section 3.4, can be thought of as a higher N_{it} shifting the probability mass to low flow states. A borewell remains functional, with positive discharge, until stochastic failure occurs with probability $\pi_{Fit} = \pi_F(N_{it}, R_{t-1})$. Since failure is an absorbing state, the failed borewell henceforth contributes nothing further to output.[19]

We assume that the probability of a successful drilling attempt, π_S, is constant. Each attempt entails a fixed cost c_d for sinking the bore hole and, if successful, an additional cost of installing a pipe, casing, and hooking up the electrical connection; the submersible pump itself is removable and thus we do not consider it a sunk cost. The total cost of a successful attempt is, therefore, $c_s > c_d$.

Finally, for the sake of tractability, we assume that at most two wells can function simultaneously on any given plot so that $N_{it} \in \{0, .., 2p_i\}$, where p_i is the number of plots in adjacency i.[20] Drilling success, failure, and discharge events for two wells on the same plot are independent random variables (*conditional* on the plot-specific unobserved heterogeneity described in Section 3.4). Using superscripts to enumerate wells, incremental output of a plot with two wells depends on the sum of their discharges $q_{it}^1 + q_{it}^2$, since water from both wells can be pooled and dispersed throughout the plot.

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

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

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

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

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

4.2 Borewell investment decision

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

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

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

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

and for the choice of making a drilling attempt

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

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

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

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

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

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

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

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

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

We can now write

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

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

Finally, we have

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

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

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

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

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

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

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. We use 14 cadastral maps representing at least one village in each mandal (see Appendix D). While the borders of these administrative maps are arbitrary in that they do not correspond to salient geographic or geological features, each contains many plots and as a result "truncation-at-border" effects should have negligible empirical consequences.[21]

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

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

Thus far, we have taken beliefs about the evolution of the number of functioning wells in the adjacency as given, deriving the plot owner's dynamic investment decision as if it were a game "against nature". However, we assume a Markov-perfect equilibrium (MPE), in which beliefs and decision rules (CCPs) of all plot owners are consistent with one another. Furthermore, our state space $(\mathcal{N}, n)$ implicitly assumes that plot owners ignore the status of wells on successive layers of plots outside their own adjacencies. This restriction is not, in general, implied by our key assumption that well interference is limited to functioning wells in the adjacency. Indeed, information on the status of wells outside the first layer might help agents predict neighbors' investment behavior and the status of wells in their

[21]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.

adjacencies, which in turn helps neighbors predict their neighbors' investment behavior and the status of their wells, and so on. Under *unrestricted* MPE play, therefore, investment decisions may depend on the state of the whole map, even with well interference effects confined to adjacent plots. To be precise, let $CCP_i(X_t)$ be a choice probability function for the owner of plot i and $\{CCP\}$ be the vector of choice probabilities of all plot owners in the cadastral map. Further, let one-period ahead transition probabilities $\widetilde{F}(X_{t+1} \mid X_t)$ describe beliefs about the evolution of the state of the map and $F(X_{t+1} \mid X_t; \{CCP\})$ be the one-period-ahead law of motion for the state induced by the primitives of the problem and $\{CCP\}$. We thus have:

Definition 1. An MPE is a vector of choice probabilities $\{CCP_i^*(X_t) : i = 1, \dots, P\}$ 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 $\widetilde{F}^*(X_{t+1} \mid X_t) = F(X_{t+1} \mid X_t; \{CCP\}^*)$.

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

As a tractable alternative, we consider a Markov equilibrium in which: i) CCPs depend only on the state of the (1-level) adjacency $(X_{(i1)t}, n_{it})$, and ii) the plot owner has beliefs only about the stochastic evolution of $X_{(i1)t}$ in steady state. While assumption i) avoids the "curse of dimensionality", the fact that well interference is largely limited to the adjacency should dampen the influences induced by unrestricted play of plot owners in layers $h > 1$ as well as making it less plausible (i.e., by bounded rationality) that plot owners would keep track of the full state of a large map. Assumption ii) is not only a natural implication of assumption i), but it also adds the non-trivial requirement that equilibrium beliefs about the state of the adjacency be correct when averaged over the map's stochastic steady state. Thus, in the spirit of an "oblivious equilibrium",[22,23] we propose

Definition 2. An Adjacency Equilibrium (AE) is a vector of choice probabilities $\{CCP_i^*(X_{(i1)t}, n_{it}) : i = 1, \dots, P\}$ and of beliefs $\left\{\widetilde{F}_i^*(X_{(i1)t+1} \mid X_{(i1)t}, n_{it}) : i = 1, \dots, P\right\}$ 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.[24] Further, let $F_i^\infty(X_{(i2)t} \mid$

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

[23] Hodgson (2021), building on the work of Fershtman and Pakes (2012) on games of asymmetric information, also uses a reduced-state equilibrium concept to obtain a tractable empirical model of strategic oil field exploration.

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

$X_{(i1)t}, n_{it}; \{CCP\})$ be the conditional distribution implied by $F^{\infty}(X_t; \{CCP\})$. Then,

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

To understand how equation (23) constrains beliefs, note first that the evolution of the state of plot $j \in \mathcal{M}_{(i1)}$ between t and $t+1$ depends on $CCP_j^*()$ at t. This investment decision rule depends, in turn, upon the state of j's adjacency at t, formed by plot j and all of its neighbors, including plot i. All of the plots in j's adjacency are in $\mathcal{M}_{(i1)}$ and $\mathcal{M}_{(i2)}$. Therefore, the state variables of plot owner j are contained in $\{n_{it}, X_{(i1)t}, X_{(i2)t}\}$. If the owner of plot i knew $X_{(i2)t}$, they would thus be able to predict their neighbor j's behavior at t using $CCP_j^*()$ and, together with other primitives such as the drilling success and well failure processes, predict the stochastic evolution of the state of plot j which is the second factor in each term of the summation in equation (23). An AE, however, assumes that $X_{(i2)t}$ in not in plot owner i's information set but that they form expectations about it using the (conditional) steady state distribution $F^{\infty}(X_{(i2)t} \mid X_{(i1)t}, n_{it}; \{CCP\}^*)$ as probability weights on the RHS of equation (23).[25] Although each plot owner still has a unique CCP and set of beliefs, and the plot owners' joint decisions still depend on the entire cadastral map, the AE concept achieves considerable simplification.[26]

5 Structural Estimation

We now detail a tractable empirical structural model and a Simulated Method of Moments (SMM) procedure to estimate it. In addition to the four structural parameters $\Omega = (\theta, \alpha, \delta, \sigma)$, Table 3 summarizes the model primitives and how they are estimated.

5.1 Empirical specification

Unobserved heterogeneity We allow for two latent plot-specific factors. The first concerns the presence/absence of water-bearing fissures in the hard-rock underlying the plot, the distribution of which we approximate as binomial; i.e., plots are either "developable" ($\mathcal{D} = 1$) for groundwater exploitation,

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

[26] 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 (SMM):		
Production function	θ, α, δ	4.1
Scale of drilling shock	σ	4.2
Fraction of developable plots	$\Pr(\mathcal{D} = 1)$	5.1
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
Unsuccessful drilling cost	c_d	note 3
Successful drilling cost	c_s	note 4
P(good monsoon|mandal)	$\mathbb{E}\, R_{mt-1}$	3.1
$P(\text{Success}\|\mathcal{D} = 1)$	π_S	2.2/note 5
Other:		
Discount factor	β	5.1
Plot network (by map-village)	$\{A, \mathbf{M}\}$	4.3

Notes: For details, see subsection in column 3 and/or numbered notes as follows: (1) Probability functions depend upon number of functioning borewells in adjacency and vary by mandal, as well as by monsoon rainfall and unobserved type; (2) Probability of (low) type 1 = 0.364, (high) type 2 = 0.636; (3) $c_d = 28,800$ Rs is median drilling cost (in 2017 Rs) across all borewells sunk since 2000; (4) $c_s = 59,800$ Rs is median of sum of drilling, pipe, casing, and electrical connection costs (in 2017 Rs) across all borewells sunk since 2000; (5) Observed drilling success per attempt = 0.375.

in which case the model of optimal annual drilling choice outlined in Section 4 applies, or not ($\mathcal{D} = 0$), in which case our model is irrelevant for that plot and there is never drilling on it. To pin down the fraction of developable plots $\Pr(\mathcal{D} = 1)$, we exploit the fraction of "active" plots (see Section 2.2) observed in our sample as discussed below. The second latent factor applies only to developable plots and concerns heterogeneity in groundwater availability from such water-bearing fissures that exist. Section 3.4 discusses estimation of the discrete distribution of this latent factor using well flow and failure data and Section 3.1 establishes its robustness to selection on developable status. We assume that plot owners observe neither the developable type nor the flow/failure type of the *other* plots in their adjacencies, which presupposes that both of these latent factors are iid across plots.[27]

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

State space restrictions and observed types We assume that CCPs depend on reference plot area a and on the number, but not on the areas, of adjacent plots. This restriction effectively reduces $X_{(i1)t}$ to $\mathcal{N}_{it} = \sum_{j \in \mathcal{M}_{(i1)t}} n_{jt}$, yielding state space $(\mathcal{N}_{it}, n_{it})$.[28] Given the one-to-one mapping between CCPs, beliefs, and developable plot types, tractability considerations also dictate limiting the number L of such types. Discretization of reference plot area into quintiles coupled with the number of adjacent plots in the maps ranging from 1 to 7, yields 35 possible observed types, which, along with 2 unobserved flow/fail types mentioned above, implies $L = 70$ for each of the 14 map villages.

Discount factor Given the challenge of identifying the discount factor in dynamic discrete choice models (Rust 1994; Magnac and Thesmar 2002), we follow standard practice by fixing the value of β in the SMM estimation. However, since our welfare calculations may be sensitive to this choice, we calibrate β: First, we estimate the structural model at different (fixed) values of β along a coarse grid. Next, we simulate from each of these estimated models in steady state the average difference of the present discounted value per acre of active plots versus inactive plots. Lastly, we compare this value differential to its empirical counterpart, which we estimate to be around 80,000 Rs/acre (see Appendix G for details), and select the β yielding the closest match. This procedure delivers a value of $\beta = 0.95$.

5.2 Solution algorithm

Given values of Ω and $\Pr(\mathcal{D} = 1)$, we obtain an AE on each of the 14 cadastral maps as follows:

Initialize the maps:

Step 1 Draw a $\mathcal{D}_j$ for each plot j from the binomial distribution with $\Pr(\mathcal{D} = 1)$.

Step 2 Assign each plot with $\mathcal{D}_j = 1$ an unobserved flow type ν_1 or ν_2, drawing from a binomial distribution with probability of (low) type $1 = 0.364$.[29]

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

Iterate on beliefs and CCPs:

Step 4 Given $\{CCP_{l,q-1} : l = 1, \ldots, L\}$ at iteration $q = 1, 2, \ldots$, 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.

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

[29] To ensure a unique AE despite the inherent randomness of a particular map draw in Steps 1 and 2, we repeat these two steps ten times for each plot type and pool the resulting data in computing beliefs in Steps 4 and 5.

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

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

Convergence:

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

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

5.3 Moment conditions and identification

To their model-based counterparts we match the observed annual drilling rates in the representative sample of plots by area quintile a_k and number of functioning borewells, empirical moments shown in panel (a) of Figure 3.[30] Since, by assumption, no investment occurs once a plot has two borewells, we do not match drilling rates conditional on $n \geq 2$. Although we do not target moments involving the number of functioning borewells outside of the reference plot (i.e., average drilling rates conditional on different values of $\mathcal{N}$), later we exploit the correlation between drilling and $\mathcal{N}$ to validate the model.

Because we estimate all parameters associated with well interference externalities in the first stage, our second-stage estimation does not require (as in the canonical case considered by Bajari et al. 2015) identifying the effect of well-owner's neighbors' actions on well-owner's payoffs. Therefore, identification of Ω may be thought of, heuristically, in terms of a static model wherein drilling decisions are made once and for all, where there is no borewell failure, and where the number of functioning borewells on a reference plot is taken as given. In this case, we have $P_{n,k} \equiv \Pr(d = 1|n, a_k) = \text{logit}^{-1}(\{\theta[f(n + 1, a_k; \alpha, \delta) - f(n, a_k; \alpha, \delta)] - \mathbb{E}\, c\}/\sigma)$, where $\mathbb{E}\, c = c_s\pi_s + c_d(1 - \pi_s)$ is the expected cost of drilling and $f = \frac{1}{\theta}\mathbb{E}[y(N, R, n, a_k; \alpha, \delta)]$ is expected (physical) output on a plot of area a_k with n functional borewells, where expectations are taken with respect to flow outcomes, beliefs about the number of functioning borewells in other adjacency plots, and monsoon rainfall.

[30]Model-based drilling rates are averages across the 14 map villages weighted by the proportion of total plot area in the representative plot sample contributed by sample plots associated with that village.

Given α and δ, the difference in drilling rates across any two (n, k) pairs with different expected physical outputs identifies the ratio θ/σ.[31] Since $\mathbb{E}\, c$ is a known constant, we can then back out σ, and hence θ, from the average drilling rate at any (n, k). Given $\theta(\alpha, \delta)$ and $\sigma(\alpha, \delta)$, the remaining eight moment conditions yield more than enough equations to solve for α and δ. Intuitively, fixing a_k, differences in drilling rates at $n = 1$ and $n = 0$ capture diminishing returns to flow because, in log odds form, $\log \frac{P_{1,k}(1-P_{0,k})}{P_{0,k}(1-P_{1,k})} = \frac{\theta(\alpha,\delta)}{\sigma(\alpha,\delta)}\{[f(2, a_k; \alpha, \delta) - f(1, a_k; \alpha, \delta)] - [f(1, a_k; \alpha, \delta) - 0]\}$. Likewise, now fixing n, differences in drilling rates across area quintiles capture how the marginal product of flow varies with area because $\log \frac{P_{n,k}(1-P_{n,k'})}{P_{n,k'}(1-P_{n,k})} = \frac{\theta(\alpha,\delta)}{\sigma(\alpha,\delta)}\{[f(n+1, a_k; \alpha, \delta) - f(n, a_k; \alpha, \delta)] - [f(n+1, a_{k'}; \alpha, \delta) - f(n, a_{k'}; \alpha, \delta)]\}$.

Finally, to identify the fraction of latent developable plots $\Pr(\mathcal{D} = 1)$, we simulate the active status of a plot of each observed type using criteria analogous to those deployed in the actual data (Section 2.2); that is, we construct synthetic 5-year panels in steady state and assign an active status indicator $\mathcal{A}$ equal to one if any drilling attempt occurs over the panel or if there is a functioning borewell in the initial period. Averaging over plots of the same area type and over 5 simulated panels per type yields the model-based moments $Pr(\mathcal{A} = 1|a_k)$ that we match to those from the representative plot sample.

In all, we have 15 moment conditions and the SMM criterion function uses a diagonal weighting matrix consisting of the inverse of these moment variances.

5.4 Results and model validation

Table 4 reports the model parameter estimates along with their bootstrapped standard errors based on 100 replications of our estimation procedure.[32] We can reject a Cobb-Douglas production function (i.e., $\delta = 0$) with a very high degree of confidence. Figures 5 and 6 also show that our model matches the 15 targeted moments reasonably well.

Table 4: Structural Parameter Estimates

θ	α	δ	σ	$\Pr(\mathcal{D} = 1)$
15.15	0.79	0.40	1.14	0.72
(0.46)	(0.02)	(0.03)	(0.04)	(-)

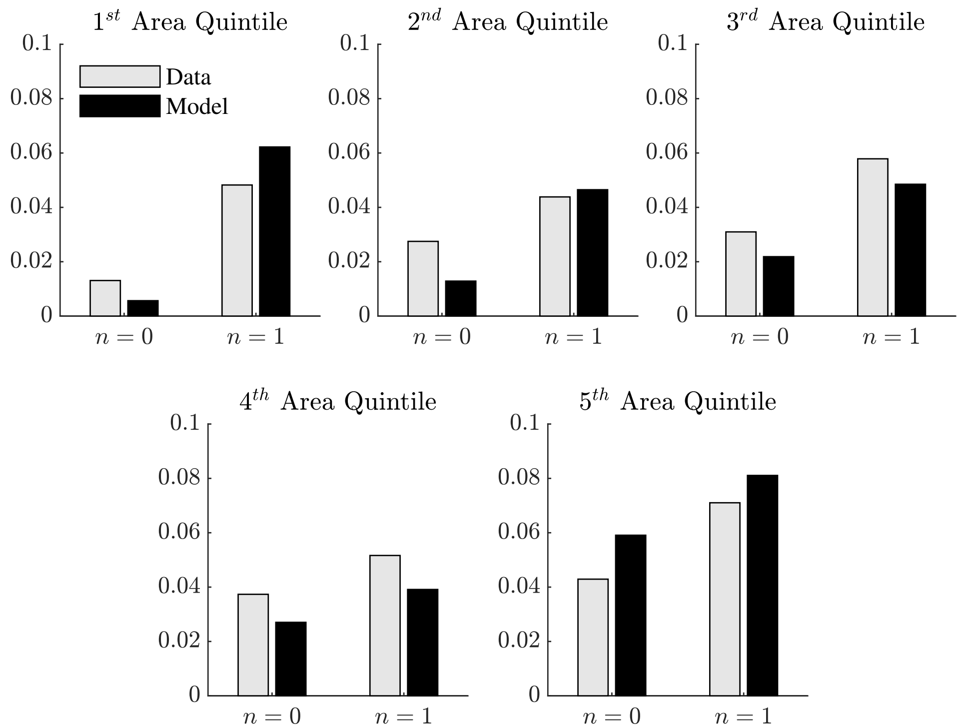
Notes: Bootstrapped standard errors in parentheses. See equation (13) for definition of production function parameters (θ, α, δ); σ is scale of drilling shock.

By way of model validation, we compute the steady state borewell density, a moment not targeted in the estimation, and obtain 0.214 wells per developable acre or 0.154 ($= 0.214 \times 0.72$) wells per total

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

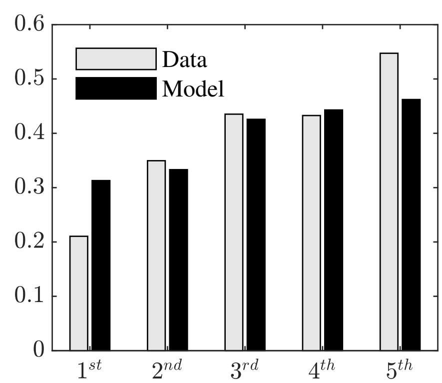
[32] These standard errors are understated because we do not account for pre-estimated first stage parameters. Given the high precision, however, any correction for first stage sampling error is unlikely to matter for inference.

Figure 5: Annual drilling rates by plot area quintile and n



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

Figure 6: Probability of developed land by plot area quintile



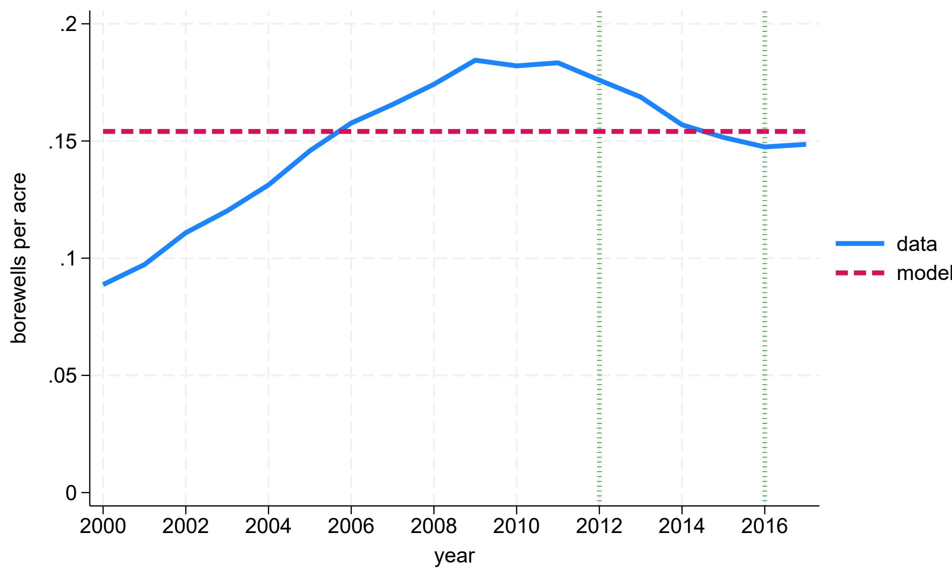
Notes: Each pair of bars represents a data moment (probability of having at least one functioning borewell or drilling attempt in a 5-year period by plot area quintile) and its corresponding model fit.

acre. Next, combining retrospective data from both the 2010 household survey and the 2017 follow-up, we obtain borewell density across our study area from 2000-2017.[33] As shown in Figure 7, density

[33]We report the 2017 figure with the caveat that number of functioning borewells in that year does not reflect failures

rose steadily throughout the first decade of the 2000s, consistent with the pattern for pre-partition Andhra Pradesh as a whole reported in Jacoby (2017). After the stock of functioning borewells peaks in 2011, a steady state appears to be reached, with density averaging 0.160 borewells per acre over the representative panel sample period 2012-16, very close to the model-generated value of 0.154.

Figure 7: Borewell density



Notes: Data are based on retrospective information from the 2010 and 2017 household surveys in Anantapur and Mahabubnagar districts. Predicted borewell density is from the model in steady state. Vertical lines demarcate the estimation sample period.

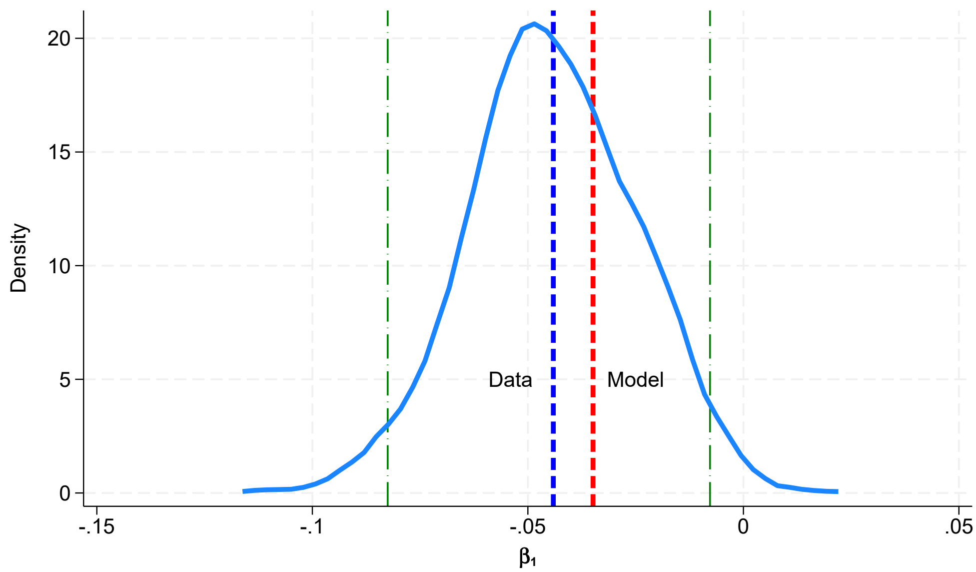
As a further validation exercise, we simulate data from the estimated model to compute the partial correlation between d_{it} and $\mathcal{N}_{it}$ and compare it to the estimated coefficient β_1 in the drilling regression (cf., equation 8 in Section 3.3). Recall that this "reduced-form" parameter captures the strength of strategic substitutability between neighbors' drilling decisions. Starting at a steady state on each of 10 replications of the cadastral village maps, we simulate five-year panels consisting of triplets $\{d_{it}, n_{it}, \mathcal{N}_{it} : t = 1, ..., 5\}$ for every plot on the map that is assigned developed status (see Step 2 in Subsection 5.3); this yields 43,702 5-year panels in total. Using this very large synthetic data "sample", we estimate a linear probability model version of the drilling reduced form (corresponding to column 5 of Appendix Table F.1), essentially giving a "population" value of β_1 from the model.[34] Figure 8 shows the bootstrap

or successful drilling attempts after May, when the 2017 household survey was administered (see Section 2.2).

[34]While we use reference plot fixed effects to account for unobserved flow/fail heterogeneity in both simulated and actual data, we do not use IV in the former case as misclassification error is not an issue. Also, since our model assumes no drilling on plots with two borewells, we drop observations with $n = 2$ in the actual and synthetic estimation samples.

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

Figure 8: Strategic interactions – data vs. model



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

6 Counterfactuals

Our quantitative policy evaluation addresses three questions: (1) What is the social cost of the current policy of free (but rationed) electricity to farmers for pumping groundwater? (2) What is the optimal tax on borewells that eliminates both the deadweight losses due to electricity subsidies and to the well interference externality? (3) How would the burden of a tax on borewells be distributed across landowners?

6.1 Social value of groundwater development

Based on our estimates, the (marginal) private value of developable land is 35,200 Rs/acre.[35,36] The social value of groundwater development, however, is this private value minus the cost of electricity, which, though given free to borewell owners, is not free to society. In particular, the steady-state fiscal cost of the electricity subsidy in present value terms is 36,400 Rs/acre = 0.214 borewells per acre $\times$ 8500/0.05, where the second term is the annual cost of electricity to run a pump given daily power rationing divided by the annual discount rate. We, therefore, find that the social value of groundwater development is essentially zero (indeed, slightly negative!) with the entire private value accounted for by the capitalized electricity subsidy.

Table 5: Electricity Cost Recovery Counterfactuals

	Network economy		*Island economy*	
	Baseline	Cost-recovery	Baseline	Cost-recovery
Borewells density (wells/acre)	0.214	0.064	0.296	0.115
Fiscal cost of subsidy ('000 Rs/acre)	36.4	0	50.3	0
Social value of GW ('000 Rs/acre)	-1.2	9.8	13.2	18.7
Deadweight loss ('000 Rs/acre)	11.0		5.5	
DWL/Fiscal cost of subsidy	0.30		0.11	

To assess the importance of the well interference externality, we compute the village map equilibria in a counterfactual economy wherein each and every plot is an island unto itself, thereby zeroing out interference effects between borewells operating in adjacent plots. As shown in Table 5, compared to the baseline "network" economy considered in the previous paragraph, we find that the steady state borewell density in the baseline "island" economy increases by 38 percent (from 0.214 to 0.296 wells/acre) and that the social value of groundwater development rises from -1,200 to 13,200 Rs/acre (private value rises from 35,200 to 63,500 Rs/acre). Thus, absent the negative externality, rent-seeking in the form of drilling for subsidies would not entirely dissipate the value of groundwater to society.

[35] This is the average expected discounted present value of agricultural output minus drilling costs in steady state across all maps weighted by the proportion of total acreage from the map village represented in the sample. Undevelopable land has zero private value by assumption (our normalization). To be clear, the private value accounts for externalities inasmuch as it averages across hundreds of adjacent plots with (potentially) mutually interfering borewells.

[36] This marginal value of developable land is much lower than the marginal value of *active* land, which we estimated at 80,000 Rs/acre, because the proportion of developable plots (0.72) is much higher than that of active plots (0.39).

6.2 Policy analysis

We now assess the social welfare implications of an annual tax τ on functioning borewells. In practice, τ could be implemented as a flat charge for maintaining an agricultural electrical connection. Setting τ for all functioning borewells equal to the annual cost of electricity (i.e., τ_e =8,500 Rs/year) would fully recover costs from agricultural consumers. A flat charge exceeding τ_e would act, at the margin, like a Pigouvian tax on borewells.[37]

We compute social welfare along the entire transition path from the zero tax baseline to the given counterfactual tax (see Appendix H for details on the equilibrium concept and solution algorithm). As in, e.g., Domeij and Heathcote (2004), this calculation takes into account the "short-run", over which the existing stock of borewells is relevant. We assume that extant borewells, if they become unprofitable, are dismantled at zero cost.[38]

Electricity cost recovery Arguing that charging farmers by electricity usage is impractical in India, Shah et al. (2007) propose flat-fee power pricing combined with quantity rationing (the latter as discussed by Ryan and Sudarshan 2022). Effectuating this policy by levying an annual tax $\tau = \tau_e$ on all functioning borewells would increase the social value of groundwater development from -1,200 to 9,800 Rs/acre. The deadweight loss from free electricity provision is thus 11,000 Rs (170 US$) per developable acre in present value terms, which is 30% of the fiscal cost of the subsidy (see Table 5). In other words, three in ten rupees transferred in-kind to farmers in the form of electricity is lost through over-drilling. By contrast, in the absence of well interference, as represented by our island economy, the fiscal cost of the subsidy is 38% higher (simply reflecting the higher equilibrium well density) but the deadweight loss is only half as large as in the network economy. Thus, the presence of externalities nearly triples the ratio of deadweight loss to fiscal cost, exacerbating the leaky bucket of free electricity provision. Despite its obvious impracticality as policy, the island economy counterfactual allows us to isolate the importance of incorporating externalities in quantifying the impacts of electricity tariff reform.

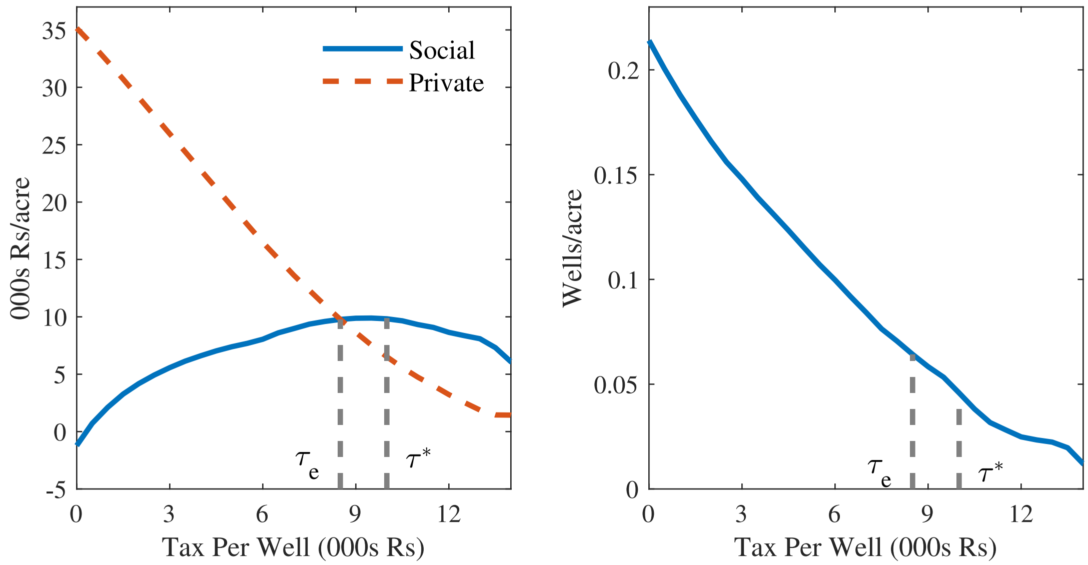
Optimal borewell taxation To correct the negative externality, the social welfare maximizing tax τ^* must exceed the annual cost of electricity τ_e.[39] In the left panel of Figure 9, we plot the social value

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

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

[39] Optimality of a centralized Pigouvian tax presupposes that landowners cannot restrain socially undesirable drilling through side-payments to neighbors. Aside from enforcement issues, one argument against such a Coaseian solution in

Figure 9: Welfare and wells under alternative borewell taxes



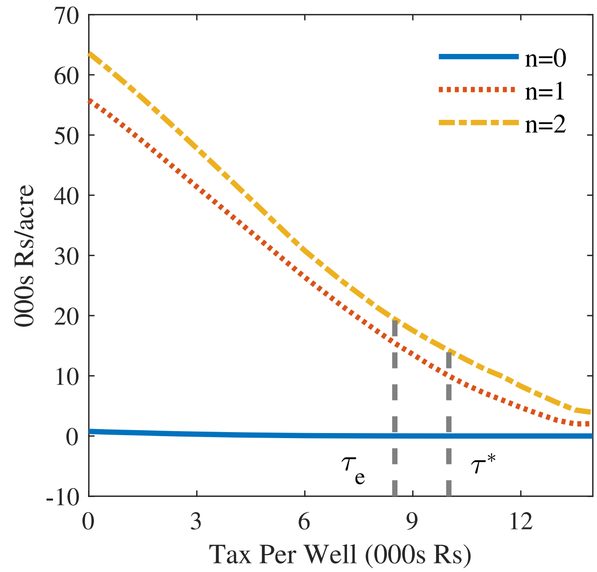
Notes: In the left panel, each point on the solid (dashed) curve represents the social (private) per acre value of developable land, or welfare, along the transition path from the benchmark zero-tax economy to the long-run steady state under alternative hypothetical annual taxes on borewells: $\tau_e = 8.5$ is the tax that recovers electricity costs; $\tau^* = 10.0$ is the optimal tax. In the right panel, each point on the curve represents the wells per acre in the the long-run steady state under alternative borewell taxes.

of groundwater development at different levels of the borewell tax and find the optimum at $\tau^* = 10,000$ Rs, about 18% higher than τ_e. Nevertheless, there is little difference in social value between a tax of 8,500 Rs and one of 10,000 Rs, and only a small absolute difference in equilibrium borewell density (right panel). Once electricity is provided to farmers at cost, reducing borewell density by 70%, the marginal externality cost is rather small.

Distributional implications of borewell taxation Although the existing stock of capital can be safely ignored when considering the very long-run effects of policy, short-run considerations are often paramount in practical policy discussions. Incorporating the short-run, as we do in this paper by accounting for transitional dynamics, pays dividends when it comes to understanding how the burden of borewell taxation would be distributed across landowners. At each counterfactual tax τ, Figure 10 shows the average private value of groundwater development per acre by the number of functioning borewells on the plot at baseline (n); the private value curve in Figure 9 (left panel) is a weighted average of these three conditional curves. Note that taxing borewells, regardless of τ, has practically zero impact on plots with no functioning borewells today because the option value of future drilling is

our setting is the complex, multi-lateral, negotiations it would require across the entire network of agricultural plots.

Figure 10: Private value by initial number of functioning borewells



Notes: Each point represents the private welfare along the transition path from the benchmark zero-tax economy to the long-run steady state under a hypothetical tax on borewells for different values of n at the time the policy is introduced: $\tau_e = 8.5$ is the tax that recovers electricity costs; $\tau^* = 10.0$ is the optimal tax.

minimal; these developable plots are heavily selected to be of the low flow/high failure type.

Turn now to the cost-recovery tax τ_e. Table 6 displays the private value of (developable) plots with, respectively, 0, 1, and 2 functioning borewells at baseline and under this counterfactual. As noted above, the welfare loss for $n = 0$ plots is negligible, whereas owners of $n = 1$ plots lose 129,000 Rs and owners of $n = 2$ plots lose 212,000 Rs. While there is considerable plot type heterogeneity *within* each of the $n = 0, 1, 2$ groups, this heterogeneity is costly, if not impossible in the case of the two latent factors, for the policy-maker to observe. By contrast, n is virtually costless to observe because a functioning borewell requires a working electrical connection. Our analysis thus suggests the contours of a practical compensation scheme to mitigate political opposition to electricity cost recovery. Suppose that each farmer with an existing agricultural connection, in agreeing to pay τ_e annually to maintain it, receives a lump-sum (per connection) of 120,000 Rs; those who successfully drill and establish a connection ex-post would not be compensated at all. The last column of Table 6 shows relatively small welfare losses or even gains under this compensation formula. Moreover, the ratio of compensation to fiscal cost of the subsidy is only 0.71 (= 120/(8.5/.05)), so the government would save nearly 30% on its budget for electricity provision while leaving farmers largely whole.[40]

[40]Although this result is merely a restatement of our earlier finding that the deadweight loss from free electricity is 30% of the fiscal cost, the efficacy of a compensation scheme based on n alone is surprising.

Table 6: Compensation Scheme for Electricity Cost Recovery

	Private value				
Plot type	Baseline	Cost-recovery	Δ	Compensation	Net welfare
$n = 0$	2.1	0.1	-2.0	0	-2.0
$n = 1$	178.6	48.3	-129.3	120.0	-9.3
$n = 2$	305.3	93.1	-212.2	240.0	27.8

Notes: All figures are in thousands of Rs per plot.

7 Conclusion

We set out to assess the social cost of a policy of free electricity to farmers for groundwater pumping in South India, a context with economically important yet highly localized externalities. To do so, we developed a tractable dynamic strategic equilibrium model of borewell investment across a large network of heterogeneous agricultural plots along with a novel simulation-based estimation strategy, one potentially applicable to a wide range of settings beyond that of concern in this paper.

Despite daily power rationing that obviates over-pumping (as per Ryan and Sudarshan 2022), once the extensive margin and drilling costs are taken into account, we find that subsidizing electricity is a rather inefficient means of transferring resources to farmers; nearly a third of the transfer value is wasted through over-drilling of borewells, a distortion greatly exacerbated by well interference. While the optimal Pigouvian tax on borewells exceeds the annual cost of electricity by some 18%, the marginal social gain from such a tax above the cost-recovery equilibrium is minimal, reflecting the already high dismantling of unproductive borewells and choking-off of drilling that such a policy would entail. Finally, although charging farmers fully for electricity would have substantial short-run distributional implications, these could largely be ameliorated through a (practical) compensation scheme.

A limitation of the empirical approach taken in this paper is that it relies on a steady state assumption; it may, therefore, not carry over to settings in which water tables are undergoing significant decline, as in much of northwest India. Adapting our methodology to these circumstances would be a fruitful area of future research.

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.

Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov (2015): "Identification and efficient semiparametric estimation of a dynamic discrete game," Tech. rep., National Bureau of Economic Research.

Benkard, C. L., P. Jeziorski, and G. Y. Weintraub (2015): "Oblivious equilibrium for concentrated industries," *The RAND Journal of Economics*, 46, 671–708.

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

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

Cole, S., X. Giné, J. Tobacman, P. Topalova, R. Townsend, and J. Vickery (2013): "Barriers to household risk management: Evidence from India," *American Economic Journal: Applied Economics*, 5, 104–35.

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

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

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

Fershtman, C. and A. Pakes (2012): "Dynamic games with asymmetric information: A framework for empirical work," *The Quarterly Journal of Economics*, 127, 1611–1661.

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.

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.

Hodgson, C. (2021): "Information Externalities, Free Riding, and Optimal Exploration in the UK Oil Industry," *unpublished manuscript*.

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

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.

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.

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

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

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

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

RYAN, N. AND A. SUDARSHAN (2022): "Rationing the commons," *Journal of Political Economy*, 130, 210–257.

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

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

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

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

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

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

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

WEINTRAUB, G. Y., C. L. BENKARD, AND B. VAN ROY (2008): "Markov perfect industry dynamics with many firms," *Econometrica*, 76, 1375–1411.

WOOLDRIDGE, J. M. (2010): *Econometric analysis of cross section and panel data.*

——— (2019): "Correlated random effects models with unbalanced panels," *Journal of Econometrics*, 211, 137–150.

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

Appendix (Not for Publication)

Contents

A Additional figures and tables **42**

B Wealth, liquidity and drilling **44**

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

D Cadastral map villages **46**

E Well failure and duration dependence **48**

F Linear estimates **49**

G Land values and active status **51**

H Transitional dynamics **53**

A Additional figures and tables

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

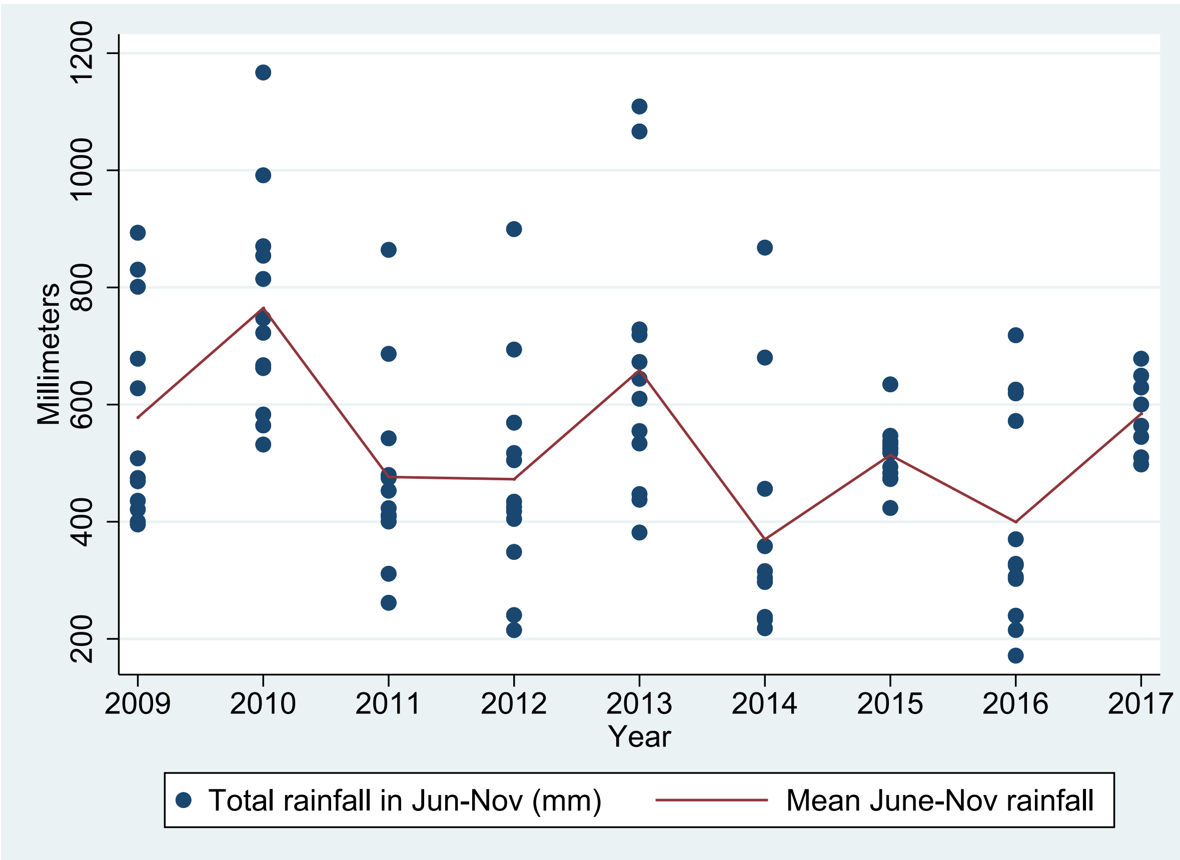


Table A.1: Well Failure by Year

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

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

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

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

B Wealth, liquidity and drilling

Using the 2010 household survey, we compute the total value of household assets in 2009, including agricultural land, livestock, agricultural machinery, household durable goods, and savings in the form of bank deposits, cash and jewelry. Gross wealth should be a good proxy for liquidity and thus allows us to test for the importance of financial constraints on well-drilling. Table B.1 presents estimates of the determinants of drilling attempts using the 2012-16 representative plot panel. Columns 1-3 report linear probability models with standard errors clustered at the plot level whereas columns 4-6 report plot-level random effects logit models. Inferences about the association between pre-sample wealth and drilling over the 5-year panel from the two sets of estimates are virtually identical.

In column 1 (and 4), which includes only mandal dummies as additional controls, we see that wealthier households (in 2009) are significantly more likely to drill during 2012-2016. However, once we also control for plot size with area quintile dummies (cols. 2 and 5), this effect largely disappears. Evidently, wealth and plot area are positively correlated and there is a greater incentive to drill on larger plots. Finally, in column 3 (and 6), we control for the initial number of functioning borewells on the plot entering the panel period, which further attenuates the wealth effect toward zero. Admittedly, these liquidity effects, or lack thereof, are associations, i.e., not necessarily causal. For example, despite controlling for initial borewells, unobserved (to us) suitability for groundwater development (and thus drilling) may have augmented pre-sample wealth through past groundwater use. Such correlations, however, cannot explain the zero liquidity effects that we find in Table B.1.

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

Table B.1: Determinants of drilling

	LPM			RE Logit		
	(1)	(2)	(3)	(4)	(5)	(6)
Log(gross wealth 2009)	0.00531***	0.00236	0.00130	0.161***	0.0701	0.0379
	(0.00189)	(0.00193)	(0.00191)	(0.0578)	(0.0622)	(0.0622)
Plot area quintile (1st omitted):						
2nd	—	0.0110**	0.00960**	—	0.563***	0.519***
		(0.00442)	(0.00444)		(0.197)	(0.197)
3rd	—	0.0210***	0.0185***	—	0.847***	0.768***
		(0.00556)	(0.00551)		(0.203)	(0.203)
4th	—	0.0222***	0.0199***	—	0.877***	0.808***
		(0.00518)	(0.00520)		(0.195)	(0.194)
5th	—	0.0284***	0.0234***	—	1.024***	0.904***
		(0.00645)	(0.00655)		(0.213)	(0.213)
Initial borewells (0 omitted):						
1	—	—	0.0197***	—	—	0.549***
			(0.00471)			(0.117)
2	—	—	0.0433***	—	—	0.859***
			(0.0125)			(0.189)
Observations	14,310	14,310	14,310	14,310	14,310	14,310
Plots	2,862	2,862	2,862	2,862	2,862	2,862

Notes: Standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$), clustered at plot-level (columns 1-3). Column 1-3 use a linear probability model; columns 4-6 use plot-level random effects logit MLE. Dependent variable is whether a drilling attempt was made on plot during the year (from 2012-16). Each regression also controls for mandal dummies (coefficients not reported).

C Average annual electricity costs per borewell

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

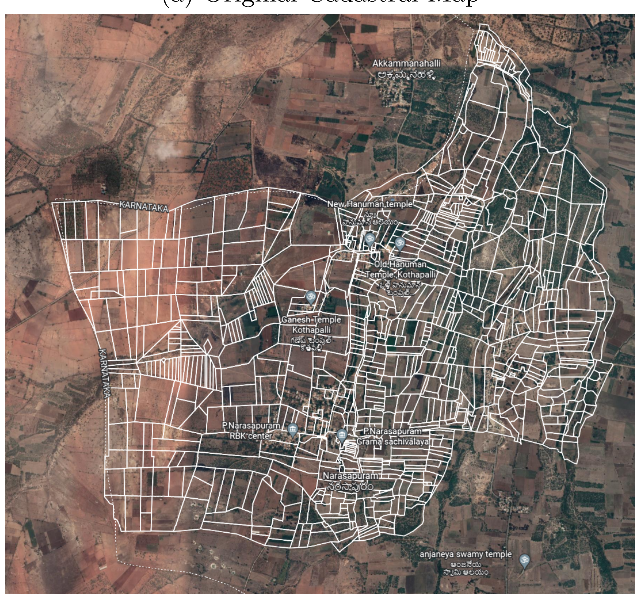
D Cadastral map villages

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

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

Figure D.1: Village Muddiredipalle

(a) Original Cadastral Map



(b) Digitized Cadastral Map

E Well failure and duration dependence

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

Table E.1: Well Age and Subsequent Failure

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

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

F Linear estimates

In this section, we present linear estimates of the drilling, flow, and failure models with reference plot fixed effects (as opposed to correlated random effects).

F.1 Drilling propensity

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

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

now treating α_i as a fixed (rather than random) effect. We assume measurement error of the classical variety and use the number of existing wells in the adjacency (outside the reference plot), $\mathcal{N}_{it}^E$, as an instrument.

Column (1) of Table F.1 reports the least-squares (FE) coefficient ignoring measurement error. As in Table 1 of the main text, we do not find a significant impact of neighboring wells on drilling. Column (2) shows the first stage regression of $\mathcal{N}_{it}$ on the instrument $\mathcal{N}_{it}^E$ and column (3) the resulting FE-IV estimate. Just as with the CRE-MEM estimator in Table 1, here we find a significantly negative effect of neighboring wells once we correct for measurement error. One concern, however, is that, if there is spatial correlation in the unobservables, then $\mathcal{N}_{it}^E$ may be correlated with the residuals, which contain the effect of *own* borewells on drilling. To assess this, in column (4) we add dummies for the number of borewells on the reference plot to remove the effect of own borewells from the residuals.[1] 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.

F.2 Well flow/failure

Instead of an ordered logit for the five flow states, we run a linear model with the dependent variable taking on values from 1 to 5 (col. 1 of Table F.2). Since N in this case is not subject to recall error, we do not instrument it. In the case of failure, we estimate a linear probability model by OLS (col. 2) and by 2SLS (col. 3) using $\log(\mathcal{N}^E + n)$ as an instrument for $\log(\mathcal{N} + n)$. Estimation samples for the separate flow and failure models are identical to those used in the joint nonlinear estimation in the main text. Results for the two sets of procedures are qualitatively similar.

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

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

	(1) drill	(2) $\mathcal{N}$	(3) drill	(4) drill	(5) drill
No. func wells exc ref plot ($\mathcal{N}$)	-0.0150		-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
Number of ref. plots	1,057	1,057	1,057	1,057	988

Notes: Standard errors in parentheses clustered by reference plot (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$). Columns 1 and 2 are by ordinary least-squares; 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.

Table F.2: Determinants of well flow and failure–Linear models

	flow (1-5)	failure (0/1)	
	(1)	(2)	(3)
log(N)	-0.513***	0.00368	0.283***
	(0.106)	(0.0523)	(0.0509)
Good monsoon	1.110***	0.00881	0.00749
	(0.347)	(0.0118)	(0.0129)
Ref plot FE	YES	YES	YES
Observations	1,028	2,851	2,851
Number of ref. plots	514	697	697

Notes: Standard errors in parentheses clustered by reference plot (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$). Columns 1 and 2 are by ordinary least-squares; columns 3 is by two stage least squares using the log number of existing wells in adjacency as an instrument.

G Land values and active status

We use plot value data collected in the 2017 household survey to estimate the difference in present discounted values between active and inactive land. Recall that an active plot is one on which at least one drilling attempt was made during 2012-16 or which already had at least one functioning borewell in 2012. Our survey asked each plot owner "if you were to sell this plot today, including the associated water rights, how much would you receive in thousands of Rs per acre?" In evaluating the present discounted value of the projected income flows off of their land, as reflected in their stated sales price, we presume that farmers use the same discount factor that they would use in assessing the future net benefits from drilling. This presumption is the basis for our calibration of β for the structural estimation.

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

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

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

Finally, we note a threat to the validity of our household fixed effect estimator: unobserved *plot-level*

Table G.1: Plot values and active status

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

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

characteristics (e.g., soil quality) correlated with both land values and active status. Recall, however, that location-specific unobservables, a far more important component of residual variation across villages than across household plots *within* villages, have little impact on our regression results. This finding suggests that any bias due to unobserved plot-level characteristics is likely to be negligible.

H Transitional dynamics

H.1 Equilibrium

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

Definition: Let $F_0^\infty(X)$ (or F_0^∞ in short) be the stationary distribution over the state of the map at the initial Adjacency Equilibrium AE$_0$. An Adjacency Equilibrium over the transition path is a vector of choice probability functions $\{CCP_{it}(X_{(i1t)}, n_{it})\}_{t=1}^\infty$ and of beliefs $\{\tilde{F}_{it}^*(X_{(i1)t+1}|X_{(i1t)}, n_{it})\}_{t=1}^\infty$ such that: a) CCPs and beliefs converge to the CCPs and beliefs of AE$_\tau$; b) given beliefs $\tilde{F}_{it}^*$, the decision rule CCP_{it}^* is the solution of the plot owner's i's dynamic game "against nature" at every t; and c) beliefs at each t are correct on "average". That is, let $F_t(X_t; \{CCP_s\}_{s=1}^t, F_0^\infty)$ be the joint distribution over the state induced by the primitives, the vector of CCPs from date $t = 1$ to t and the initial steady state distribution of the map F_0^∞. Further, let $F_t(X_{(i2)t}|X_{(i1)t}, n_{it}\{CCP_s\}_{s=1}^t, F_0^\infty)$ be the conditional distribution implied by $F_t(X_t; \{CCP_s\}_{s=1}^t, F_0^\infty)$. Then,

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

H.2 Solution algorithm

Recall that in our empirical structural model we reduce the dimensionality of the AE by partitioning the set of adjacencies into types, such that all adjacencies of the same type share beliefs and CCPs, and we limit the state of the adjacency to $(\mathcal{N}, n)$. We compute an AE along the transition path as follows:

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

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

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

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

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

Step 4.1 Start in period T-1.

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

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

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

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

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