Investment Booms and Natural Resources *

Pedro Gete[†], Michael Reher[‡], and Athena Tsouderou[§]

September 2026

Abstract

Investment booms create an environmental externality through endogenous producer composition: who invests, not just how much, determines the damage to shared natural resources. We show this using the mid-2010s boom in California tree nuts, where short-term investors with little prior industry exposure drove nearly all net cropland expansion, irrigating with groundwater subject to a common pool problem. Areas more exposed to these investors saw significantly greater groundwater declines and well drilling. This channel accounts for roughly one-fifth of the groundwater decline in agricultural California from 2013 to 2023. A simple framework shows the mechanism is not more capital chasing higher profits, which would predict conservative extraction, but entry by low-quality, short-horizon producers with little incentive to conserve a resource they will not depend on. The lesson generalizes to booms drawing on any common-pool input: composition, not just scale, governs the external cost.

Keywords: Investment booms, commodities, cropland, entry, investors, groundwater, water, well drilling.

JEL Classification: G11, G23, D84, Q15, Q25, Q54.

*We thank seminar participants at University of Cambridge, Chinese University of Hong Kong, University of California San Diego, University of Navarra, AREUEA, IESE, Hoyt Institute, Big Sky Conference and Almond Alliance's Annual Convention for comments. Yaru Su and Haoran Wang provided excellent research assistance. Research Reported in this paper was partially funded by MICIU /AEI /10.13039/501100011033 / FEDER, UE, Grant No. PID2024-161318NB-I00

[†]IE University. Email: pedro.gete@ie.edu.

[‡]University of California San Diego, Rady School of Management. Email: mreher@ucsd.edu

[§]University of Miami, Herbert Business School. Email: atsouderou@miami.edu.

1 Introduction

Booms are usually judged by how much capital gets built. This paper shows that who builds it matters just as much, because the composition of the investors a boom attracts determines the use of the shared natural resource. Investment often follows a boom-bust pattern (e.g., Hayek 1931; Kydland and Prescott 1982), and it is well known that private investment during a boom can generate pecuniary externalities or otherwise induce spillovers to the real economy (e.g., Geanakoplos and Polemarchakis 1985; Greenwald and Stiglitz 1986). This paper empirically shows that investment booms can also amplify environmental externalities. This is fundamentally a composition story: the damage does not follow mechanically from the amount of capital a boom builds, but from who builds it.

Our laboratory is the mid-2010s boom in California tree nuts and the associated depletion of groundwater, the resource used to irrigate the trees. This setting is well suited to our question for four reasons, detailed in Section 3. Groundwater scarcity is a first order global problem, with documented links to migration and unrest elsewhere (e.g., Jasechko et al. 2024; Inamda 2025; Shokri 2026), and California groundwater has declined steadily for decades. Tree nut planting is a long duration, water intensive investment that depends on groundwater once surface water is rationed during drought, making the resource rival and non excludable. Tree nut investment follows a recurring boom-bust cycle, and the mid-2010s episode was driven by a run up in prices that attracted new capital. Critically, nearly all of the net expansion in nut cropland during this boom came from late entrants, investors with little prior exposure to the industry, who expanded on cheap, low quality land and behaved in ways inconsistent with a long term horizon.[1]

Using parcel level data on ownership, land cover, water consumption, and well drilling across agricultural California over 2013-2023 (Section 2), we compare changes in groundwater levels in stations with more versus less expansion by late entrants (Section 4). Stations more exposed to late entrants experience significantly greater declines in groundwater levels and

[1] As an anecdotal example, the largest late entrant in our sample, a private equity fund named Trinitas Partners, sold off its portfolio as the boom ended. Popular narratives support our empirical findings, with reports in local newspapers alleging that Trinitas drilled groundwater wells to irrigate almond orchards planted on former pastureland, while stating: "We [Trinitas] didn't really know much about the almond industry when we started ... But we liked the fact there's a growing worldwide demand for almonds, and even when prices are at normal levels, you can make a decent return" (Sbranti 2014).

significantly more well drilling (Section 5), a result that holds after controlling for land cover, soil, drought exposure, pre trends, and existing growers' own expansion, and that strengthens under 2SLS estimates instrumenting for late entrants' expansion with proximity to nut processors (Section 6).

A natural concern is that this is simply the mechanical consequence of more total investment: more cropland, regardless of who plants it, depletes more groundwater. Our evidence rules this out. If late entrants merely added to total investment, standard channels would predict them extracting groundwater *less* intensively per unit of capital, not more. Overly optimistic expectations about future nut prices can explain why late entrants invest more overall, but such optimism would incentivize them to extract less groundwater per tree, to avoid raising their own future extraction costs; less efficient production technology implies the same, since groundwater is a complement to other inputs. Instead, we find that late entrants have shorter holding periods and are more likely to sell after the boom, consistent with short term preferences lowering the value of preserving the aquifer, which incentivizes more extraction today (Section 8). In short, late entrants extract more per unit of investment not because they invest more, but because of who they are.

To summarize, our results support, in order of importance, that investment booms create an environmental externality through endogenous producer composition; that late entrants have shorter horizons and extract more groundwater per unit of investment than existing producers; that this is not merely the mechanical effect of greater total investment; that the mechanism is quantitatively important, with late entrants' expansion accounting for around one fifth of the overall decline in groundwater levels over 2013-2023 in our sample of agricultural California (Section 7); and that the implication plausibly extends to other booms that draw on common or exhaustible resources, including mining, energy, fisheries, forestry, and data center development.

Put simply, the composition of investors matters as much as does the level of investment. While our setting is groundwater and tree nuts, the same logic should apply wherever capital investment competes for a shared, depletable input, such as critical minerals for battery and renewable energy buildout, fish stocks for fleet expansion, or grid and water capacity for data centers, since investors who enter late and expect to exit early have systematically weaker

incentives than long horizon investors to conserve the resource they draw on. From a policy perspective, these findings may provide a new, environmental rationale for macroprudential regulation. That said, there are two caveats. First, we cannot estimate adaptation to declining groundwater levels, which would plausibly attenuate the effects we find (e.g., Hornbeck and Keskin 2015). Second, we cannot estimate within aquifer groundwater spillovers, which implies that our results may recover a lower bound on the effects of late entrants (e.g., Ayres, Meng and Plantinga 2021).

Related Literature

First, we contribute to a broad literature on free entry by providing new empirical evidence that unrestricted entry can generate inefficiencies, especially in a real estate context. Our focus on the economic effects of late entrants in nut cropland parallels a quickly-growing literature on the economic effects of out-of-town entrants in residential real estate markets.[2] Favilukis and Van Nieuwerburgh (2021), for example, show how out-of-town homebuyers can raise prices and reduce welfare for local residents; analogously, we document how entrants to an agricultural boom impose costs on existing users by depleting groundwater. A number of other papers also study the effects of out-of-town entrants, such as Chinco and Mayer (2016), Badarinza and Ramadorai (2018), Gao, Sockin and Xiong (2020), Cvijanović and Spaenjers (2021), or Gorback and Keys (2025), but the literature varies in how, economically, it interprets these entrants. In our setting, one can interpret late entrants as similar to short-term housing speculators in DeFusco, Nathanson and Zwick (2022) or Bayer et al. (2020), as they enter the market following a rise in expected cash flow and exit shortly thereafter. A key difference, however, is that late entrants also accelerate natural resource depletion, in addition to the effects they have on real estate prices.

Second, we contribute more generally to a large macro-finance literature on recurring investment cycles. Empirically, a recent set of papers has documented such cycles in a variety of contexts, focusing on how forecasting frictions such as diagnostic expectations (e.g., Bordalo, Gennaioli and Shleifer 2018; Bordalo et al. 2020), competition neglect (e.g., Greenwood and

[2]Free entry can also be inefficient by eroding industry profits in the context of monopolistic competition, as shown theoretically by Mankiw and Whinston (1986). Other empirical contexts in which free entry can be inefficient include real estate agents (e.g., Hsieh and Moretti 2003; Barwick and Pathak 2015).

Hanson 2015), or inability to observe permanent demand shocks (e.g., Povel et al. 2016) can amplify investment during the boom. Relative to these papers, we focus on the *consequences* of recurring investment booms rather than their expectational origins, although it seems natural that any of the above forecasting frictions would amplify the effects of booms on natural resources.[3] Theoretically, a related stream of papers has characterized how private investment can generate pecuniary or aggregate demand externalities or otherwise induce capital misallocation.[4] By contrast, the externality that we emphasize − excess depletion of a common natural resource to build capital − is conceptually quite simple and distinct from pecuniary or aggregate demand externalities, although we leave open whether and how these different frictions would interact in a general setting.

Third, we relate to recent empirical papers on the economics of groundwater in California's Central Valley. We complement Hadachek et al. (2026), who also study how shocks to growers in agricultural California affect groundwater levels over recent decades. While Hadachek et al. (2026) focus on shocks originating from heat and drought, we instead focus on how shocks to expected profits from entering a boom in perennial cropland affect groundwater. In the opposite causal direction, Hagerty (2022) and Burlig, Preonas and Woerman (2026) show how making it harder to access surface water and groundwater, respectively, affect the area and composition of cropland in the long run.

2 Data

We merge from different sources to create a unique data set that tracks ownership, crop allocation, groundwater well drilling, and other variables on land parcels in California over 2013 through 2023. We also collect data on depth to the aquifer at repeat measurement stations throughout the state. Full details on our data are in Appendix A.

[3]To be clear, we are by no means the first to document recurring boom-bust cycles in cropland. Since at least Kaldor (1934), Ezekiel (1938), and other papers summarized in Chavas, Chambers and Pope (2010), it has been known that such cycles exist and pose challenges for the full-information-rational-expectations benchmark. We differ by utilizing the cross-section of investors who enter the boom to understand how such booms can affect resource extraction.

[4]A very partial set of recent papers characterizing pecuniary externalities associated with private investment includes Shleifer and Vishny (1988), Lorenzoni (2008), He and Kondor (2016), Dávila and Korinek (2018), Lanteri and Rampini (2023). A very partial list characterizing aggregate demand externalities includes Farhi and Werning (2016), Korinek and Simsek (2016), and Rognlie, Shleifer and Simsek (2018).

Cropland Data Layer

Data on crop allocation come from the United States Department of Agriculture's (USDA) Cropland Data Layer (CDL), which we simply refer to as CropScape. The CropScape data contain an annual snapshot of what covers a given unit of land, based on satellite imagery. We measure changes in land covers from 2013, the first year of the drought, through 2023. This ten-year frequency ensures that our data can capture slow-moving changes in land cover due to the fact that tree nuts require time from planting to maturity, and it roughly corresponds to the historical length of the nut price cycle. Appendix A describes the layout of the CropScape data set, how we map crop layers from this data set to land parcels, and how we compute the share of land in each land parcel allocated to different uses. To assuage concerns that the CropScape data systematically mis-classify certain crop types, we replicate our main results using land cover data from a private vendor, LandIQ (Appendix B.3).

Our cleaned CropScape data set contains the share of land allocated to a given crop in 2013 and 2023 on all land parcels within the major agricultural counties of California. This set of counties is defined by selecting counties in order of agricultural land area until we reach 95% of all cropland area in California, according to the 2022 USDA Census. Unless otherwise specified, we filter the data to exclude parcels with less than half of their area allocated to active or fallowed cropland in 2023 and to exclude parcels owned by the federal or state government, which are typically cases of eminent domain or reclaimed tribal land. Data on ownership come from the Corelogic data set, described next.

CoreLogic

We obtain data on the ownership and characteristics of agricultural properties from the CoreLogic Assessment Data Set, which aggregates tax assessment records from county assessor offices. We merge the tax assessment data from CoreLogic with the CropScape data set using assessor's parcel number and match 95% of parcels in the latter data set. We refer to this parcel-level data set as the "CoreLogic-Cropscape" data set.

To study ownership, we construct an owner identifier using a string grouping algorithm described in Appendix A.3 applied to the owner's name in the assessment data. We then

compute the total acreage in the CropScape data set associated with each owner over time as well as the share of this acreage allocated across different crops or land uses.

California Department of Water Resources

Data on groundwater measurement stations, aquifer depths, and net well drilling are from the California Department of Water Resources (DWR). For background, aquifers are underground bodies of porous rock or sediment that collect groundwater. The DWR defines groundwater basins and subbasins as geographic units with significant overlap in their aquifers. See Appendix Figure A.2 for an illustration. The DWR relies on a collection of measurement wells within each geography to monitor groundwater levels, typically expressed as the depth to the aquifer underground. These measurement wells, which we call "stations", are the most granular unit of geography at which we observe aquifer depths.

While measured aquifer depths are similar within groundwater basins, there is still variation within each basin that we could potentially use for identification. In particular, Bierkens and Wada (2019) note that aquifers do not resemble a bathtub, in that groundwater pumping does not homogeneously and uniformly affect groundwater levels within an aquifer. Therefore, in the interest of maintaining sample size, we conduct our core analysis at the level of the groundwater measurement station. We aggregate from our parcel-level data set to a station-level data set by matching each land parcel to the nearest groundwater measurement station. Each station has an average of 680 acres and represents a collection of land parcels.

In addition to collecting data on measurement stations, the DWR also records the date and location of private well drilling for non-measurement purposes, stored in its Well Completion Reports data set. We compute the total number of well-feet drilled on each parcel over 2013 through 2023. As before, we aggregate parcel-level well drilling to the level of the measurement station.

Other Data Sets

Appendix A describes auxiliary data sets used either directly in our analysis or as intermediate steps to build and merge the main data sets. These auxiliary data sets include:

cropland returns from the National Council of Real Estate Fiduciaries (NCREIF); county-level drought data from the National Drought Mitigation Center; aggregate tree nut statistics from the USDA and the Almond Board of California; land parcel shapefiles from the Los Angeles County Data Hub; groundwater basin shapefiles from the California Department of Water Resources' Bulletin 118; the location of tree nut processing plants from trade groups for almonds, pistachios, and walnuts; microdata on consumptive water use (i.e., evapotranspiration) from Open ET; and a data set on surface water allocations produced by Hagerty (2022) that derives from the DWR's eWIRMS data set and other public sources.

2.1 Summary Statistics

Our main analysis sample consists of 1,336 groundwater measurement stations summarized in Table 1. We also perform a parcel-level analysis to sharpen the mechanism, and Appendix Table A.2 summarizes the 51,200 parcels used in this analysis. Appendix A.4 provides a step-by-step description of how we arrive at this sample size through each stage of the data merging and cleaning procedure.

3 Key Facts

Our empirical setting is the mid-2010s boom in almonds, pistachios, and walnuts, which we simply refer to as tree nuts. We now describe several facts that make this setting ideal for studying how investment booms affect natural resources. Appendix D elaborates.

Planting Nut Trees is a Capital Investment. Nut trees require between three and six years from the time of planting until the trees first bear fruit (e.g., Pollack 1998). Because of this time-to-build feature, planting nut trees is a capital investment, as it is in the model of Dorfman and Heien (1989).

Tree Nut Production is Geographically Concentrated. Tree nuts require a specific rain cycle, air humidity, and soil nutrient profile, which leads to geographic concentration in certain regions, like California. In fact, California accounts for 77%, 65%, and 24% of global almond, pistachio, and walnut production, respectively, according to the USDA Foreign Agricultural Service (2025). This concentration implies that global nut demand can affect

the supply of non-tradable inputs used in California's nut production, such as water.

Nut Prices and Investment Follow Recurring Booms. Like with other capital goods, tree nut production follows an investment cycle. We walk through the key dynamics for the case of almonds. Figure 2 reveals three almond cycles over the past thirty years, all of which begin with a sharp increase in real almond prices. These sharp price increases are soon followed by steep increases in acres of young trees (i.e., non-bearing acres), which are a measure of almond tree investment. Several years after the increase in investment, the quantity of almonds produced begins to rise (Appendix Figure A.4). Our empirical analysis focuses on the boom of the mid-2010s, which also coincides with a boom in pistachios and walnuts documented in Appendix Figure A.5.

Tree Nut Production is Water-Intensive. Water is the most important input used in tree nut production (e.g., Yaghmour et al. 2016; Brar et al. 2015; Hasey et al. 2015). Relative to other crops, almond, pistachio, and walnut tree nuts require substantially more water, even when conditioning on their relatively high nutritional content per unit produced, as calculated by Fulton, Norton and Shilling (2019) and summarized in Appendix Figure A.1.

Groundwater is a Common Pool. Growers can irrigate nut trees using either surface water, which collects in bodies of water on the surface of the earth, or groundwater, which collects in underground bodies of porous rock or sediment called aquifers. Unlike surface water, groundwater in California is unregulated in the mid-2010s, and, thus, a non-excludable good. Moreover, groundwater's relatively slow rate of natural replenishment limits supply, making it a rival good. These two facts make groundwater a common pool.

Drought Incentivizes Use of Groundwater. The tree nut boom of the mid-2010s occurred amid a significant drought in California, which Lund et al. (2018) summarize. During droughts, leasing surface water can become prohibitively expensive (Appendix Figure A.3). Groundwater, by contrast, is non-traded (Appendix D). So, the primary private costs of acquiring it are well drilling and pumping, which are less drought-sensitive. Thus, nut growers have an incentive to tap groundwater for irrigation during the boom. Indeed, Figure 3 shows how land parcels with growth in nut trees account for slightly less than half of groundwater well drilling in agricultural California over 2013-2023. Interestingly, much of this drilling

occurred over 2014-2016, at the height of the boom in nut tree planting.[5]

Late Entrants Drive the Boom. Late entrants to the nut boom account for the majority of aggregate expansion of nut cropland over 2013-2023, as shown in Figure 4. We define late entrants as growers that allocated less than 10% of their California agricultural land to tree nuts in 2013 or did not own such land then. So, by definition, late entrants collectively owned only a very small share of California nut cropland in 2013. Yet, as the boom progresses, they accumulate nut cropland until owning around half of the overall stock in 2023.

Appendix Table A.1 reports the largest overall growers of nut cropland in the middle and the end of the boom, as well as the largest growers among late entrants into the boom. Interestingly, fiduciaries such as REITs (e.g., Gladstone Land Corp), providers of retirement plans (e.g., TIAA), and other investment vehicles predominantly owned by pension funds (e.g., Trinitas) feature among the largest orchard owners, as do traditional agricultural firms, such as Wonderful Nut Orchards and JG Boswell Company, which specialized in cotton before the nut boom of the mid-2010s.

Late Entrants and Groundwater. Putting the previous two facts together, we ask whether the late entrants who drive the boom could reduce the supply of California's groundwater by extracting it to irrigate their orchards. Figure 5 provides suggestive evidence consistent with this possibility. Specifically, the figure shows how measurement stations where late entrants expanded by more over 2013-2023 experience significantly greater increases in depth to the aquifer over that period.[6] Of course, this figure leaves open whether late entrants behave any differently than existing growers, as well as whether their expansion could simply proxy for other determinants of aquifer depth. The next section describes our strategy for isolating causality in Figure 5.

[5]A property's owner makes the decision to drill a well, since well drilling is a capital investment and requires the owner to sign the permit application. For this reason, we do not emphasize differences between orchards operated by the owner versus by an outside operator. According to the NCREIF data set, 74% of acres owned by fiduciaries in a given year are managed by a direct subsidiary of the owner.

[6]The scatterplot is binned, so that each marker represents around 20 stations, each of which is 680 acres. Note that, for a small subset of stations, late entrants' expansion is actually negative. This occurs when late entrants acquire nut orchards and subsequently tear down the trees. Loosely speaking, these unusual cases may be interpreted as late entrants "shorting" the boom.

4 Identification

We want to identify the effect of late entrants' tree nut investment on the change in groundwater levels in agricultural California over 2013-2023. For now, we do not attach an economic interpretation to this estimand. Instead, we devote Section 8 to a detailed analysis of what it means to be a late entrant in terms of fundamentals that would apply to other settings.

Identifying the effect of late entrants' investment through the time series would be impossible because planting decisions occur at a low frequency. So, we instead appeal to cross-sectional variation across groundwater measurement stations, s. We estimate

$$Y_s = \beta \Delta LateEntrantLandShare_s + \delta \Delta OtherNutLandShare_s + \gamma X_s + u_s, \tag{1}$$

where: $\Delta LateEntrantLandShare_s$ is the change in the share of s allocated to almond, pistachio, and walnut trees over 2013 through 2023 on land owned by late entrants midway through the boom, in 2018; and $\Delta OtherNutLandShare_s$ is the similar change in share of s allocated to the three tree nuts on land not owned by late entrants mid-boom.[7] The most important outcome Y_s is the change in depth to the aquifer over 2013-2023, in feet. Correspondingly, u_s contains unobserved recharge shocks, which raise groundwater levels, or unobserved drivers of extraction, which lower them. We conduct inference by clustering u_s at the level of the groundwater subbasin associated with a station, since water from recharge shocks induces within-subbasin correlation in aquifer depth.

The coefficient β recovers the causal effect of late entrants' expansion on Y_s if

$$0 = \mathbb{E}\left[\Delta LateEntrantLandShare_s \times u_s | \Delta OtherNutLandShare_s, X_s\right]. \tag{2}$$

Assumption (2) states that unobserved drivers of changes in groundwater levels or well drilling in the cross-section of stations are *conditionally* uncorrelated with late entrants' expansion. The conditioning arguments are important, since late entrants' decision of where

[7] Appendix Table A.8 verifies that the results are similar if we measure the middle of the boom as 2021. Since the USDA Cropscape data set calculates land covers at the pixel level (i.e., every 900 square meters), any measurement error in the share of a station allocated to tree nuts will be larger for small stations. We address this issue by weighting stations by their land area (Solon, Haider and Wooldridge 2015). For the same reason, we weight parcels by land area when estimating the parcel-level regressions in Section 6.

to expand is clearly non-random. Perhaps the most important conditioning argument is $\Delta OtherNutLandShare_s$, which absorbs a variety of factors related to tree nut planting as reflected in the expansion of existing growers. We further control for a host of variables: initial land covers and depth to the aquifer in 2013; soil characteristics used in the literature, including soil pH, water storage, and K(w) factor (e.g., Deschênes and Greenstone 2007); contemporaneous 2013-2023 changes in different types of non-nut land covers, where non-cropland is the residual cover; and measures of drought severity and exposure to future groundwater regulations associated with the Sustainable Groundwater Management Act, which Bruno and Hagerty (2025) find can incentivize anticipatory groundwater extraction.

There are three main ways in which assumption (2) could potentially be violated. The first concerns reverse causality: late entrants, knowing that they will rely on groundwater for irrigation, choose to avoid stations where groundwater levels are expected to fall. Importantly, such reverse causality would bias the estimated effect towards zero and, so, if anything, would make the results conservative. The second potential violation is an omitted variable not included in the previous set that jointly affects changes in depth to the aquifer and late entrants' expansion. We shall address this concern in a number of ways, such as controlling for and inspecting pre-trends, using an instrumental variable related to ease of tree nut processing, and estimating variants of equation (1) using a parcel-level analysis that exploits only within-station variation. A third potential violation of assumption (2) is purely spurious correlation, which we address by replicating our results across a wide variety of specifications, data sets, and geographic units.

5 Effect on Aquifer Depth

Table 2 contains the estimates of equation (1) when the outcome is well-drilling, and Table 3 contains the estimates when the outcome is change in depth to the aquifer. While this latter outcome is our ultimate interest, we first describe Table 2 because the way through which late entrants could affect aquifer depths is through pumping groundwater wells they drill.

Effect on Well Drilling. Column (2) of Table 2 implies that a standard deviation increase in $\Delta LateEntrantLandShare_s$ leads to a statistically-significant 0.2 standard deviations more

well-feet drilled. The insignificant coefficient on $\Delta OtherNutLandShare_s$ suggests that late entrants drill at a higher intensity when they expand than do existing nut growers, and we investigate why this is the case in Section 6.3. Column (3) verifies that the results are robust to controlling for well drilling over 2008-2012, a pre-period.

Effect on Aquifer Depth. Column (2) of Table 3 implies that stations with more expansion by late entrants experience a larger increase in depth to the aquifer, that is, a larger decline in groundwater levels. Given the results in Table 2, this pattern likely reflects how late entrants drill at a higher intensity, which would enable them to extract more groundwater using the drilled wells. Appendix Table A.12 obtains similar results when specifying the outcome as the log change in depth to the aquifer, although, from a hydrologic perspective, specifying the change in levels as in Table 3 is more correct (e.g., Scanlon et al. 2002).

Recalling assumption (2), the fact that the coefficient on $\Delta LateEntrantLandShare_s$ nearly doubles while the R^2 increases when moving from the uncontrolled specification in column (1) to the controlled one in column (2) suggests that unobserved omitted variables would bias the estimate towards zero (Oster 2019), setting aside the case of heterogeneous treatment effects that we consider in Appendix B.5. This increase in the coefficient is also consistent with the remark from Section 4 that late entrants have an incentive to avoid expanding in areas with declining groundwater levels. In columns (3)-(5), we again find similar results after controlling for pre-trends in the change in depth to the aquifer. Appendix Table A.12 further shows that late entrants' expansion does not predict the change in depth to the aquifer over these various pre-periods, which serves as a placebo test and, thus, supports the validity of the main effect.[8]

Overall, the results in Table 3 document a robust cross-sectional relation between late entrants' nut expansion and declines in groundwater levels. Several pieces of evidence just described suggest that the relation is causal, and the next section performs additional tests that further support this causal interpretation.[9]

[8]This finding also provides evidence against differential pre-trends where late entrants expand, although we cannot perform a proper event study for two reasons. First, changes in groundwater levels occur over relatively low frequencies, and so the existence or non-existence of year-to-year variation in groundwater levels may be spurious. Second, as mentioned in Appendix A.4, only a subset of wells provide measurements for consecutive years over long horizons and especially for periods before 2013, which is why our sample size falls in columns (3)-(5) of Table 3.

[9]Appendix Table A.6 separately estimates the effect of $\Delta LateEntrantLandShare_s$ from land that late

6 Internal Validity

6.1 Where Late Entrants Expand

To better understand how the main effect is identified, we regress $\Delta LateEntrantLandShare_s$ on a large set of pre-determined variables. We standardize the variables to ease interpretation. The results are in column (1) of Table 4.

Highlighting some of the correlations: late entrants expand in stations with a larger initial depth to the aquifer, more erodible soil, a smaller initial share of nut cropland, and a larger initial share of non-cropland, which mostly consists of open space or pastureland. That late entrants expand on land with a larger initial aquifer depth suggests that our results do not confound drilling incentives of unobserved groundwater users. Such users have an incentive to drill where the initial depth is smaller, which is not where late entrants expand. Moreover, that late entrants expand on non-cropland suggests that, in a counterfactual without the nut boom, this land may have experienced less cultivation and, thus, less well drilling.

At an intuitive level, several of these correlations suggest that late entrants expand on cheap land. We investigate this interpretation through a hedonic regression of log sale prices on these same station-level characteristics, based on a sample of real estate transactions for parcels in our main data set that occur over 2013-2023, excluding purchases made by late entrants. Using the results of this hedonic regression, which is summarized in column (3) of Table 4, we then calculate the fitted values at the parcel-level and aggregate them to the station level. Lastly, we regress $\Delta LateEntrantLandShare_s$ on these fitted values, which we interpret as land quality.

The results in column (2) of Table 4 indeed show that late entrants expand more where the market value of land is cheap, based on these ex-ante characteristics. This interpretation likely does not reflect a deliberate attempt to acquire lower-quality land, but, rather, to avoid incurring fixed costs associated with buying expensive land.

entrants already owned in 2013 versus from land that they acquired, finding a much stronger effect from the former. This may reflect how late entrants who are able to acquire land endogenously choose to expand in areas with less of a drop in groundwater levels, again reinforcing the interpretation of the baseline estimates as conservative. However, the fact that 65% of late entrants' expansion occurs on land that they already owned suggests that illiquidity in the agricultural real estate market constrains their choice of land on which to expand.

Recapping, Table 4 shows that late entrants expand in areas where, plausibly, there would have been less well drilling without the boom (i.e., on non-cropland with a larger initial depth to the aquifer) and where land values are cheap, based on a hedonic quality adjustment. This latter pattern of expanding on lower-quality land is consistent with late entrants not intending to stay in the market for the long-term.

6.2 Instrumental Variable

Next, we limit the identifying variation to a single source, namely, proximity to technologies that are specific to the harvesting and processing of tree nuts. We re-estimate equation (1) through 2SLS using the average distance to the nearest nut processor among parcels in station s, denoted $DistanceToProcessor_s$, as an instrument for $\Delta LateEntrantLandShare_s$. Importantly, we exclude the 8% of processing plants built after 2014 to avoid reverse causality from processors endogenously locating themselves nearby the orchards of late entrants.

We have already seen in column (1) of Table 4 that $DistanceToProcessor_s$ strongly predicts where late entrants expand, which supports its relevance as an instrument. At the same time, column (3) shows how this variable is uncorrelated with sale price for real estate transactions not involving a late entrant, which suggests that proximity to nut processors is not valuable for reasons apart from more efficient nut production, thus supporting the exclusion restriction.

The statistical relevance of being close to a nut processor makes practical sense given the importance of drying out tree nuts after harvest (e.g., U.S. Environmental Protection Agency 2025) and the risk of introducing mold or aflatoxins if the drying is not performed (Chen and Pan 2022). Being close to a processor further improves production because seasonal labor and harvesting technology (e.g., shakers) can congregate in the vicinity of a processor. It also seems plausible that

$$0 = \mathbb{E}\left[DistanceToProcessor_s \times u_s | \Delta OtherNutLandShare_s, X_s\right], \tag{3}$$

which means that, conditional on the variables in our main regression, proximity to a processor only affects well drilling and groundwater levels through its effect on late entrants' nut

expansion. Assumption (3) is much weaker than its baseline counterpart, assumption (2).

Column (1) of Table 5 reports the estimates of the following equation,

$$\Delta LateEntrantLandShare_s = \zeta DistanceToProcessor_s + \tilde{\delta}\Delta OtherNutLandShare_s + \tilde{\gamma}X_s + u_s. \quad (4)$$

The negative and highly significant estimate of ζ support the relevance of average distance to the nearest processor as an instrument. However, to limit weak-instruments bias in the second stage, we exploit heterogeneity in the effect of $DistanceToProcessor_s$ on $\Delta LateEntrantLandShare_s$, and we include interactions between the control variables X_s and $DistanceToProcessor_s$ in the vector of instruments (Abadie, Gu and Shen 2024).

Columns (3) and (5) of Table 5 report the results of the corresponding second-stage regression when the outcomes are well-feet drilled and change in depth to the aquifer, respectively. The estimates are roughly twice as large as their OLS analogues in columns (2) and (4), and remain significant after applying the Lee et al. (2022) standard error correction. On the one hand, this larger estimated effect may reflect how late entrants have an incentive to avoid expanding in areas with declining groundwater levels, thus biasing the OLS estimates towards zero as discussed in Section 4. On the other hand, the larger estimates could simply reflect weak instruments bias. However, the first stage F-statistic of 33.5 lies above the Stock and Yogo (2005) cutoff for 5% maximal bias, suggesting that any such bias is not severe. The insignificant J-statistic does not reject exogeneity of the entire instrument set.

Recapping, the results of this instrumental variable analysis suggest that the baseline estimates in Tables 2 and 3 may be conservative.

6.3 Within-Station Analysis

To further avoid bias from unobserved station-level characteristics, we perform a parcel-level analysis that limits identifying variation to the same station through the inclusion of station fixed effects. We focus on well drilling as our outcome of interest, since changes in depth to the aquifer are only observed at the station level, and, thus, would be absorbed by the

station fixed effect. Specifically, we estimate

$$WellFeetPerAcre_{i,\text{13-23}} = \beta LateEntrant_{o(i)} + \gamma X_i + \alpha_{s(i)} + u_i, \tag{5}$$

where $LateEntrant_{o(i)}$ indicates if the owner o of parcel i midway through the boom (i.e., in 2018) allocated less than 10% of their total land to tree nuts in 2013 and, thus, is a late entrant. The sample is restricted to parcels with growth in nut cropland.[10]

The results in Table 6 imply that late entrants drill more well-feet-per-acre than existing growers over 2013-2023. The robustness to controlling for well-drilling during the pre-period shows that this finding does not mechanically reflect how existing growers may have already drilled their wells before the boom, per column (3) of Table 6.

These parcel-level results strongly support the validity of the baseline station-level results, as they hold fixed all station-level unobservable characteristics. Yet, these results also raise the question of how existing nut growers irrigate their land if they do not rely as much on well drilling as nearby late entrants within the same station. We investigate two possible margins of adjustment in Table 7.

Surface Water Rights. One margin of adjustment is the source of water, holding total water applied fixed. In particular, late entrants may have a higher well drilling intensity than existing growers because they do not own land with rights to surface water. Unlike groundwater, surface water rights are regulated by the California DWR, although they are almost never curtailed (Hagerty 2022). Moreover, when traded, surface water rights command a high price in times of drought (Appendix Figure A.3). Together, these observations imply that surface water rights are a valuable asset. The costly process of acquiring land with such rights is a fixed cost that protects the owner from elevated water prices during droughts and the associated incentive to drill groundwater. In column (2) of Table 7, we find that late entrants are less likely to own land with surface water rights.[11] So, part of the reason why

[10]The inclusion of station fixed effects $\alpha_{s(i)}$ means that we must also limit our set of controls X_i, to those observed at the parcel-level, namely land cover shares in 2013 and contemporaneous changes in non-nut land cover over 2013-2023. As noted in footnote 7, parcels are weighted by land area. Since we do not study groundwater levels as the outcome, we simply calculate heteroskedasticity robust standard errors.

[11]We restrict attention to post-2014 surface water rights, which are subject to annual reporting requirements and, thus, more precisely measured.

the depth to the aquifer increases by less where existing growers expand in Table 3 may be because these existing growers are more likely to have land with surface water rights. So, they do not need to rely as much on groundwater.

Water Consumption. A second margin is the total quantity of water applied. While we do not observe water applied at the parcel level, we do observe evapotranspiration, which is a commonly used measure of water consumption (e.g., Allen et al. 1998; Boyce et al. 2020). As Appendix A describes in detail, evapotranspiration is calculated from satellite measures of vapor emitted by the earth and by plants, which we access through the platform Open ET. Differences between evapotranspiration and water applied can arise due to differences in water storage, runoff, or percolation rates, which are unlikely to vary substantially within a given station and which scale multiplicatively with total water applied. Hence, we specify the outcome as log of annual evapotranspiration over 2013-2023. The results in column (3) of Table 7 imply that late entrants consume more water than existing growers. Quantitatively, the difference is a modest 1.3 log points, or around 5.2 liters per square meter per year (Appendix Table A.12). So, the fact that late entrants consume more water is unlikely to fully explain why their expansion corresponds to a greater decline in groundwater levels, as suggested by the back-of-envelope calculation in Appendix B.4. Instead, a combination of both relying relatively more on groundwater than surface water and consuming more water overall seems to best explain why the depth to the aquifer increases more in stations where late entrants expand more than existing growers.

6.4 Additional Robustness.

We conduct a variety of other robustness tests described in detail in Appendix B. Here, we briefly summarize the main findings. The general pattern across these tests aligns closely with the main results from Section 5, supporting their internal validity.

6.4.1 Errors in Name Grouping Algorithm

Our name grouping algorithm might overstate the number of late entrants insofar as it fails to assign the same owner identifier to variant spellings of what is essentially the same entity.

As we clarify formally in Appendix B.1, such under-grouping would only explain our results if owners with names more prone to under-grouping systematically own land in stations with increasing aquifer depths over our analysis period. While such a relation seems unlikely, we nevertheless proceed by restricting our analysis to owners whose names are plausibly less prone to under-grouping, finding similar results as also described in Appendix B.1.

6.4.2 Sensitivity to Measurement Stations as Unit of Analysis

In Section B.2, we replicate our results at a different unit of analysis, namely, at the level of detailed-analysis-unit-by-counties (DAUCOs) as defined by the California DWR. We obtain similar results as in Table 3, despite a significant drop in sample size. This similarity implies that our main finding is not driven by how we define measurement stations.

6.4.3 Errors in Satellite Land Cover Classifications.

We replicate our results using data from a private vendor, LandIQ, that relies on proprietary data to distinguish between different types of nut trees. As we describe in Appendix B.3, the results are very similar to the baseline findings. This similarity implies that the baseline results are not driven by measurement error in the USDA's satellite classification of land cover.

6.4.4 Weights of Heterogeneous Treatment Effects

In Appendix B.5, we allow for heterogeneous effects of late entrants' expansion across the controls in equation (1). This exercise accounts for the possibility that adding controls to a regression changes the weights across treatment effects that comprise the coefficient of interest, β (e.g., Angrist 1998). The results are similar to those in Tables 2 and Table 3.

6.4.5 Tree Density and Water Requirements by Tree Age

In Appendix Table A.7, we find similar results as in Table 6 after normalizing parcel-level well drilling by different denominators. First, we normalize by acres of nut cropland − not total acreage − to account for the possibility that late entrants drill more per-acre because

they plant nut trees more densely and with fewer cover crops. Second, we normalize by the growth in tree nut acreage over 2013-2023. We interpret this other normalization as well drilling per acre of young trees, which require less water than mature trees (e.g., Yaghmour et al. 2016; Brar et al. 2015; Hasey et al. 2015).

7 Aggregation

We use the estimates from Table 3 to calculate how much higher aggregate groundwater levels would have been without investment in tree nuts by late entrants. We focus on an in-sample effect, so that "aggregate" may be interpreted as privately-owned agricultural land in California, primarily the Central Valley. This exercise requires three more assumptions.

1. **General Equilibrium.** Existing owners of nut cropland would have replaced a share λ of the tree nut investment lost when late entrants are removed.

2. **Aquifer Geometry.** The average depth to the aquifer within each station equals the depth to the aquifer at the measurement well.

3. **Spillover Effects.** Removing late entrants from the overall market would not affect the depth to the aquifer in stations where late entrants did not actually expand.

Assumption 1 recognizes that, if consumers' nut demand is partially inelastic, then existing owners would expand by more in a counterfactual without late entrants than they do in the data. In particular, without late entrants, future nut supply would be lower, future nut prices would be higher, and existing owners' incentive to invest in tree nuts would thus be stronger. Of course, the share of late entrants' investment that would have been replaced by existing owners in general equilibrium also depends on a number of other difficult-to-calibrate parameters. For this reason, we simply provide bounds on the aggregate effect by assuming the existing owners replace a share $\lambda \in \{0, 1\}$ of the lost investment.

Assumption 2 implies that we can obtain an aggregate effect by weighting measurement stations by their land area. Since we do not observe the geometry of the aquifer, it is possible that the depth to the aquifer at the exact point of the measurement well may under

or overstate the average level of water within the surrounding geography that comprises the "station". However, the assumption is less inaccurate for stations that are contained within a single groundwater subbasin, defined as a geographic unit that shares the same aquifer. Reassuringly, this condition applies to 95% of stations in our estimation sample.

Assumption 3 reflects how, depending on the rate of hydraulic conductivity across stations, late entrants' expansion can affect aquifer depths even in stations where they do not expand. However, our research design cannot identify these spillover effects. So, we simply assume that they are zero. Ayres, Meng and Plantinga (2021) confront a similar issue in the context of a spatial-discontinuity design, and they show how the inability to identify groundwater spillovers produces a lower bound.

Under Assumption 1, we can compute the aquifer depth under a counterfactual without late entrants in station s as

$$\begin{aligned} CounterfactualDepth_{s,23} = -\beta_\lambda \Delta LateEntrantLandShare_s + \\ \Delta AquiferDepth_s + AquiferDepth_{s,13} \end{aligned} \tag{6}$$

where: $\beta_\lambda = \beta$ when $\lambda = 0$; $\beta_\lambda = \beta - \delta$ when $\lambda = 1$; and, from equation (1), β is the coefficient on $\Delta LateEntrantLandShare_s$ and δ is the coefficient on $\Delta OtherNutLandShare_s$. Using Assumptions 2 and 3, we can write the aggregate change in the volume of groundwater under the counterfactual as

$$\begin{aligned} CounterfactualGroundwaterLoss_{13\text{-}23} = \sum_s \big[CounterfactualDepth_{s,23} - \\ AquiferDepth_{s,13}\big] \times Area_s, \end{aligned} \tag{7}$$

where $Area_s$ is the area of station s, in acres. By contrast, the actual groundwater loss is

$$ActualGroundwaterLoss_{13\text{-}23} = \sum_s \Delta AquiferDepth_s \times Area_s, \tag{8}$$

We are particularly interested in the share of the actual groundwater loss that would not have occurred if there had not been an investment boom with late entrants. We denote this

statistic as η, which equals

$$\eta = \frac{ActualGroundwaterLoss_{13\text{-}23} - CounterfactualGroundwaterLoss_{13\text{-}23}}{ActualGroundwaterLoss_{13\text{-}23}} \tag{9}$$

Table 8 reports the result of this exercise. The top row shows the actual increase in aquifer depth of 6.11 feet-per-acre, as already reported in Table 1. The second row reports the counterfactual increase in aquifer depth from equation (8). Column (1) reports a counterfactual increase in aquifer depth of 5.1 feet-per-acre when $\lambda = 1$. The counterfactual equals 2.6 when $\lambda = 0$, shown in column (2). Mechanically, the counterfactual increase in aquifer depth is smaller when $\lambda = 0$ because existing owners do not replace the nut tree investment of late entrants. The fact the counterfactual increase is still less than the actual loss when $\lambda = 1$ reflects how expansion by existing owners raises aquifer depths by less than expansion by new owners: $\beta > \delta$. In light of Tables 6 and 7, this inequality reflects how existing owners drill with less intensity, are more likely to have acquired land with surface water rights, and consume less water.

Turning to the bottom row, the share of aggregate groundwater loss due to late entrants, η, lies between 17% and 57%, depending on the value of λ. We err on the side of the lower value given evidence that tree nut demand is inelastic (Dharmasena and Capps Jr 2017) or has near-unit elasticity (Bakhtavoryan et al. 2022). So, we conclude that investment during the nut boom by late entrants accounts for around one-fifth of the decline in groundwater levels. We reach a similar conclusion when performing the same exercise but using estimates from the LandIQ data and from a log-linear specification (Appendix Table A.9).

Assessing Externalities via Property Price Returns.

The previous calculation does not have welfare content and, in particular, does not attempt to quantify externalities associated with late entrants' expansion. In Appendix Table A.11, we move towards assessing such externalities by estimating the effect of late entrants' expansion on property price returns of other landowners in the same station who sell over 2024-2025, after the boom. These other landowners experience a 1.8 pp lower annualized return in stations where late entrants expand by an additional standard deviation, consistent with late

entrants exacerbating a common-pool externality by depleting groundwater in the boom.

8 Mechanisms

We turn to the economic mechanisms that could explain late entrants' behavior, and, in particular, why they appear to extract groundwater at a higher rate than existing growers. We organize our analysis around the following stylized framework. We stress that the framework is only intended to guide our analysis, not to deliver general theoretical results.

There are three time periods, $t \in \{0, 1, 2\}$. Time 0 corresponds to the start of the boom. Growers invest in nut trees that begin to bear fruit at time 1, as the boom peaks. The trees bear fruit again at time 2, which we interpret as a compound period that collects time after the boom. Since our main empirical results in Table 3 focus on differences in groundwater extraction for a given level of tree planting, we hold the level of investment, and, thus, the stock of trees, fixed at K for $t \geq 1$.

Growers use groundwater w_t to irrigate the trees, which produce output Y_t according to $Y_t = A(w_t)K$. The function $A(\cdot)$ describes the orchard's yield, which we assume to be increasing in w_t within a neighborhood of the problem's optimum. The yield may also depend on unmodeled management practices. Extracting groundwater comes with the proportional cost $c(D_t)$, where D_t is the depth to the aquifer. The function $c(\cdot)$ is increasing and convex in D_t, consistent with Gisser and Sanchez (1980) or Burlig, Preonas and Woerman (2026).

We now introduce subscript i to index orchards. The owner of orchard i may also own other symmetric orchards that, collectively, comprise a share ϕ of the land overlying the aquifer. The grower's competitors own the remaining $1 - \phi$. Groundwater extracted from orchard i increases the depth to the aquifer under other orchards according to hydraulic conductivity, θ. So, depth to the aquifer under orchard i evolves as

$$D_{i,t+1} - D_{i,t} = w_{i,t} + \theta \left[\phi w_{i,t} + \int_{\phi}^{1} w_{j \neq i,t} dj \right] - z_{i,t}, \tag{10}$$

where $z_{i,t}$ is a recharge shock (e.g., rainfall). Equation (10) implies that growers recognize how extracting groundwater increases their own future pumping costs − and more so if ϕ

has positive measure − but they do not internalize their effect on other growers because groundwater is a common-pool resource. This gives rise to a drawdown externality.[12]

After planting the trees, the expected value of orchard i as of time 0 is

$$V = \max_{\{w_{i,t}\}} \sum_{t=1}^{t=2} e^{-rt} \mathbb{E}\left[p_t A(w_{i,t})K - w_{i,t}c(D_{i,t})\right], \tag{11}$$

where p_t denotes the stochastic price of nuts, and r denotes discount rate. Growers solve problem (11) subject to equation (10). The first order condition with respect to $w_{i,1}$ is

$$0 = p_1 A'(w_{i,1})K - c(D_{i,1}) - e^{-r}\mathbb{E}_1\left[w_{i,2}c'(D_{i,2})(1+\phi\theta)\right]. \tag{12}$$

Condition (12) provides important economic intuition about candidate mechanisms that could explain our empirical results. We collect this intuition in the following remark, verified in Appendix C, and then we discuss the evidence for and against each mechanism.

Remark 1. Let $w^*_{i,1}$ denote the solution to condition (12). Then $w^*_{i,1}$ is: increasing in r ("Preference for Short-Term Cash Flows"); decreasing in p_2 ("Expected Persistence of the Boom"); decreasing in ϕ when ϕ is small ("Ownership of the Common Pool"); and increasing in a multiplicative increase in $A(\cdot)$ at time 1 ("Technology and Experience").

8.1 Preference for Short-Term Cash Flows

A preference for the short term may make late entrants willing to exchange higher future costs of extraction for higher short-term cash flow. By contrast, existing growers may limit short-term irrigation to avoid raising their own future extraction costs. Such a short-term preference could arise through a number of economic channels. In the literal context of the framework, it arises because of a higher discount rate, which lowers the present value of future extraction costs: $\partial w^*_{i,1}/\partial r > 0$. More generally, a preference for the short-term can

[12] In keeping with the framework's emphasis on economic intuition, equation (10) abstracts from the cone of depression, which implies that within-aquifer flows depend on within-aquifer elevation differences, as described in detail by Edwards (2016). In addition, there are a number of other groundwater externalities besides the drawdown externality that we do not account for, such as the erosion of a buffer to offset future recharge shocks (e.g., Edwards and Guilfoos 2021).

also arise through other channels, such as a shorter holding period. Having a shorter holding period incentivizes extraction at time 1 if extraction reduces property prices by less than it reduces the present discounted value of nut profits, as we show formally in Appendix C.

Since we cannot observe discount rates, we assess the role of short-term preferences by testing whether late entrants have shorter holding periods. We study three measures. The first is an indicator for whether a grower sells their parcel during the bust period of 2024-2025. The second measure is the realized holding period for parcels that were nut orchards in 2013 and were sold over 2024-2025, which follows the approach in Bayer et al. (2020) and DeFusco, Nathanson and Zwick (2022).[13] The third measure is the potential holding period for parcels that were nut orchards in 2013 and hypothetically sell in 2024. Using a parcel-level specification similar to that in Section 6.3, we then regress each of these measures on an indicator for whether the owner of parcel i in the last year before the bust is a late entrant.

The results in column (1) of Table 9 show that parcels owned by late entrants at the end of the boom are significantly more likely to be sold during the bust than parcels owned by existing growers. Moreover, the results in columns (2)-(3) verify that late entrants have shorter realized and potential holding periods, respectively. Completing the logic, column (4) shows how parcels with a shorter potential holding period as of 2024 experience significantly greater well drilling over 2013-2023. Specifying the independent variable as a potential holding period avoids reverse causality from the fact that realized holding periods depend on actions taken during the boom.

Collectively, these findings are consistent with the average late entrant having a greater preference for short-term cash flows than the average existing grower, leading late entrants to extract groundwater more intensively.

Financial Distress. A caveat to this interpretation is that we cannot separate holding periods that were intended to be short from holding periods that were inadvertently shortened due to financial distress. For example, if motivations unrelated to short-term preferences lead

[13]Restricting to parcels that sell over 2024-2025 ensures that the corresponding holding period overlaps with the tree nut boom of 2013-2023. Restricting to parcels that were nut orchards in 2013 avoids cases where the owner converts non-nut cropland to nut orchards over the boom, so that all ownership spells in the sample have a similar interpretation as buy-and-maintain investments as of the start of the boom. We define nut orchards as parcels with at least 10% of their area allocated to tree nuts in 2013. Because our CoreLogic data end in 2023, we obtain data on sales over 2024-2024 from a similar data provider, Regrid, as described in Appendix A.

late entrants to deplete the aquifer to the point of reducing their own cash flow or property values, then they may default on their creditors and sell their properties in distress. Indeed, the average late entrant who sells over 2024-2025 does so at a nominal capital loss of 2.2% annualized, although this loss is indistinguishable from zero (Appendix Figure A.6).

While entering financial distress at a higher rate may partly explain why late entrants are more likely to sell in the bust, this explanation must wrestle with two observations. First, it cannot explain why shorter *potential* holding periods correspond to more well drilling in Table 9, whereas a preference for short-term cash flows can. Second, the financial distress interpretation would likely need to appeal to strategic motives, since elevated net income over the boom would reduce the pressure to default for lack of liquidity. As a benchmark, nominal annualized income returns on the portfolios of institutional investors in the NCREIF data set were 8.4% and 11.4% for almonds and pistachios, respectively, over 2013-2023. So, if financial distress drives columns (1)-(2) of Table 9, then it must either be because late entrants earn income returns far below the benchmark, or because they strategically default at a higher rate for reasons unrelated to short-term preferences.

8.2 Ownership of the Common Pool

Another possibility raised by Remark 1 is that late entrants own smaller shares of the common pool, which reduces their incentives to preserve it: $\partial w^*_{i,1}/\partial\phi < 0$. We test this possibility by constructing two measures of common pool ownership. The first is grower o's share of land overlying the groundwater measurement station s as of the midpoint of the boom (2018). The second measure is defined similarly in terms of o's share of land overlying the subbasin $b(s)$. We estimate

$$ShareOfGeographyOwned_{o,g} = \beta LateEntrant_o + \alpha_g + u_{o,g}, \tag{13}$$

where the geography g is either the station or the subbasin.

The results in columns (1) and (3) of Table 10 show that late entrants own smaller shares of the stations or subbasins in which they operate, relative to other landowners. Columns (2) and (4) then verify that owning smaller shares of a geography corresponds to more average

well drilling in that geography over 2013-2023. Qualitatively, these results might suggest that existing growers extract less than late entrants because they own more geographically-concentrated orchards; consequently, extracting more on one parcel raises their extraction costs at their other, nearby parcels. Quantitatively, however, this mechanism seems weak. For example, multiplying the coefficients in columns (1) and (2) implies that late entrants' geographic dispersion raises their extraction by only 0.002 well-feet-per-acre, and a similar value obtains from multiplying columns (3) and (4). These quantities lie far below the far below the estimates in Table 6, although we acknowledge that the specifications also differ.

Overall, differences in ownership of the common pool seem to contribute somewhat to why late entrants extract at a higher rate, even if this mechanism does not drive the effect.

8.3 Expected Persistence of the Boom

The recurrent tree nut booms documented in Figure 2 and Appendix Figure A.5 suggest that growers overreact to changes in nut prices − in the ex-post sense − like firms do more generally in other investment settings (e.g., Bordalo, Gennaioli and Shleifer 2018; Bordalo et al. 2020; Greenwood and Hanson 2015; Povel et al. 2016). Indeed, the anecdotal evidence and present-value regressions in Appendix B.6 supports this view that the aggregate nut market over-estimates the persistence of booms. The fact that late entrants drive the recent nut boom might suggest that they may have particularly optimistic expectations about the boom's persistence.[14]

However, expectations of a longer boom would actually incentivize growers to extract *less* groundwater during the boom so that they preserve the aquifer for sustained higher prices in the future. The stylized framework makes this point clear with the result $\partial w^*_{i,1}/\partial \mathbb{E}_1[p_2] < 0$. So, even if late entrants do have more optimistic expectations, these expectations would not explain why they extract groundwater at a higher rate. Instead, having more optimistic expectations could explain why late entrants expand more in the first place, which can be seen in our framework by endogenizing K (Appendix C).

[14]Growers cannot lock-in future nut prices because there are no futures markets for tree nuts over our sample period. Neither does the USDA's Risk Management Agency offer price insurance for tree nuts, although it does offer highly standardized yield (i.e., quantity) insurance. See Appendix D for details.

8.4 Technology and Experience

Technology. If late entrants have an inferior production technology, A, then they may only enter the market in booms, when income is high enough to break even. Our finding that late entrants expand on cheap land supports this view (Table 4). However, tree nut production is well-approximated by Leontief technology (Dorfman and Heien 1989). This means that groundwater complements other inputs, such as soil quality, fertilizer, or pollination, as captured by a multiplicative increase in $A(\cdot)$ in Remark 1. Consequently, having an inferior technology would *discourage* groundwater extraction, since groundwater cannot substitute for technological inefficiency. So, differences in technology between late entrants and existing growers seems unlikely to explain why the former extract at a higher rate.

Experience. Another possibility is that late entrants extract at a higher rate simply out of ignorance. For example, they may not realize that groundwater extraction actually raises the depth to the aquifer, or they − or the farm operators they employ − may irrigate more than is optimal. While we do not observe the knowledge base of late entrants, the ample availability of public growing manuals and resources from commodity boards makes pure ignorance an unlikely explanation (e.g., Yaghmour et al. 2016; Brar et al. 2015; Hasey et al. 2015). To nevertheless evaluate the role of inexperience, we re-estimate equation (1) after allowing separate effects by whether the late entrant owned agricultural land in California in 2013, and, thus, has experience in agriculture. Appendix Table A.6 reports similar coefficients regardless of experience, suggesting that experience does not explain our results.

8.5 Summary of Mechanisms

Collectively, our empirical results are consistent with a model where investment booms attract short-term producers who extract resources at a higher rate than existing producers. While we have focused on why late entrants extract more groundwater per unit of nut cropland, we have not tried to explain why they plant more nut cropland overall. One possibility is that late entrants have an inferior production technology (e.g., Table 4). So, they only enter the market during booms, when elevated nut prices enable them to break even. Since late entrants no longer find it profitable to remain in the market during the bust, they leave,

which effectively gives them a preference for short-term cash flows. Completing the logic, this effective short-termism would then incentivize higher rates of extraction during the boom.

9 Conclusion

We document an important spillover of real investment booms: by attracting short-term investors with higher rates of resource extraction, investment booms can accelerate natural resource depletion. Our setting is the mid-2010s boom in California tree nuts and associated depletion of groundwater to cultivate the trees. A spike in expected profits from tree nut investment attracts late entrants to the market. This leads to a boom in real investment and well drilling to tap groundwater for irrigation. Relative to existing growers, late entrants have a shorter holding period, own a smaller share of the underlying aquifer, and are more likely to expand on cheap land with a higher pre-existing depth to the aquifer. Aggregating the estimates from our cross-sectional regressions implies that late entrants to the tree nut boom account for around one-fifth of the decline in aquifer levels in our sample.

From a policy perspective, our results may provide a novel rationale for macroprudential policies: excessive private investment during booms may not only generate pecuniary or aggregate demand externalities, but they may also amplify environmental externalities through the depletion of scarce natural resources used to build or to complement capital. Future research may characterize how these externalities interact in equilibrium and whether optimal policy would differ after accounting for the natural resource costs of investment booms.

More directly, our findings raise the question of whether and how policymakers should regulate groundwater. In California specifically, local jurisdictions have struggled to establish a market for groundwater largely due to steep transaction costs (Ayres, Edwards and Libecap 2018). So, policymakers at the state level passed the 2014 Sustainable Groundwater Management Act (SGMA), which requires local jurisdictions to construct and implement plans for groundwater management by the early 2040s. While we do not study SGMA specifically, our results suggest that the composition and, in particular, the holding period of groundwater users would be a relevant input for the construction of groundwater management plans.

References

Abadie, A., Gu, J. and Shen, S.: 2024, Instrumental variable estimation with first-stage heterogeneity, *Journal of Econometrics* **240**(2), 105425.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M.: 1998, *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*, FAO Irrigation and Drainage Paper No. 56, Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.

Angrist, J. D.: 1998, Estimating the labor market impact of voluntary military service using social security data on military applicants, *Econometrica* **66**(2), 249–288.

Ayres, A. B., Edwards, E. C. and Libecap, G. D.: 2018, How transaction costs obstruct collective action: The case of california's groundwater, *Journal of Environmental Economics and Management* **91**, 46–65.

Ayres, A. B., Meng, K. C. and Plantinga, A. J.: 2021, Do environmental markets improve on open access? evidence from california groundwater rights, *Journal of Political Economy* **129**(10), 2817–2860.

Badarinza, C. and Ramadorai, T.: 2018, Home away from home? Foreign demand and London house prices, *Journal of Financial Economics* **130**(3), 532–555.

Bakhtavoryan, R., Cheng, G., Capps, O. and Dharmasena, S.: 2022, A household-level demand system analysis of nuts in the united states, *Agricultural and Resource Economics Review* **51**, 283–310.

Barwick, P. J. and Pathak, P. A.: 2015, The costs of free entry: an empirical study of real estate agents in greater boston, *The RAND Journal of Economics* **46**(1), 103–145.

Bayer, P., Geissler, C., Mangum, K. and Roberts, J. W.: 2020, Speculators and middlemen: The strategy and performance of investors in the housing market, *The Review of Financial Studies* **33**(11), 5212–5247.

Bierkens, M. F. and Wada, Y.: 2019, Non-renewable groundwater use and groundwater depletion: a review, *Environmental Research Letters* **14**(6), 063002.

Bordalo, P., Gennaioli, N., Ma, Y. and Shleifer, A.: 2020, Overreaction in macroeconomic expectations, *American Economic Review* **110**(9), 2748–2782.

Bordalo, P., Gennaioli, N. and Shleifer, A.: 2018, Diagnostic expectations and credit cycles, *The Journal of Finance* **73**(1), 199–227.

Boyce, S. E., Hanson, R. T., Ferguson, I., Schmid, W., Henson, W. R., Reimann, T., Mehl, S. W. and Earll, M. M.: 2020, One-water hydrologic flow model: A modflow based conjunctive-use simulation software, *Technical report*, US Geological Survey.

Brar, G. S., Doll, D., Ferguson, L., Fichtner, E., Kallsen, C. E., Beede, R. H., Klonsky, K., Tumber, K. P., Anderson, N. and Stewart, D.: 2015, Sample costs to establish an orchard and produce pistachios.

Bruno, E. M. and Hagerty, N.: 2025, Anticipatory effects of regulating the commons, *Journal of Environmental Economics and Management* p. 103183.

Burlig, F., Preonas, L. and Woerman, M.: 2026, Groundwater and crop choice in the short and long run, *Technical report*.

Campbell, J. Y. and Shiller, R. J.: 1989, The dividend-price ratio and expectations of future dividends and discount factors, *The Review of Financial Studies* **1**(3), 195–228.
URL: *https://academic.oup.com/rfs/article/1/3/195/1580239*

Chavas, J.-P., Chambers, R. G. and Pope, R. D.: 2010, Production economics and farm management: A century of contributions, *American Journal of Agricultural Economics* **92**(2), 356–375.
URL: *https://onlinelibrary.wiley.com/doi/10.1093/ajae/aaq004*

Chen, C. and Pan, Z.: 2022, Postharvest processing of tree nuts: Current status and future prospectsâa comprehensive review, *Comprehensive Reviews in Food Science and Food Safety* **21**(2), 1702–1731.

Chinco, A. and Mayer, C.: 2016, Misinformed speculators and mispricing in the housing market, *The Review of Financial Studies* **29**(2), 486–522.

Cvijanović, D. and Spaenjers, C.: 2021, "We'll always have Paris": Out-of-country buyers in the housing market, *Management Science* **67**(7), 4120–4138.

Dávila, E. and Korinek, A.: 2018, Pecuniary externalities in economies with financial frictions, *The Review of Economic Studies* **85**(1), 352–395.

DeFusco, A. A., Nathanson, C. G. and Zwick, E.: 2022, Speculative dynamics of prices and volume, *Journal of Financial Economics* **146**(1), 205–229.

Deschênes, O. and Greenstone, M.: 2007, The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather, *American economic review* **97**(1), 354–385.

Dharmasena, S. and Capps Jr, O.: 2017, Consumer demand for nut products in the united states: Application of semi-parametric estimation of censored quadratic almost ideal demand system (c-quaids) with household-level micro data.

Dorfman, J. H. and Heien, D.: 1989, The effects of uncertainty and adjustment costs on investment in the almond industry, *The Review of Economics and Statistics* pp. 263–274.

Edwards, E. C.: 2016, What lies beneath? aquifer heterogeneity and the economics of groundwater management, *Journal of the Association of Environmental and Resource Economists* **3**(2), 453–491.

Edwards, E. C. and Guilfoos, T.: 2021, The economics of groundwater governance institutions across the globe, *Applied Economic Perspectives and Policy* **43**(4), 1571–1594.

Ezekiel, M.: 1938, The cobweb theorem, *Quarterly Journal of Economics* **52**(2), 255–280.

Farhi, E. and Werning, I.: 2016, A theory of macroprudential policies in the presence of nominal rigidities, *Econometrica* **84**(5), 1645–1704.

Favilukis, J. and Van Nieuwerburgh, S.: 2021, Out-of-town home buyers and city welfare, *The Journal of Finance* **76**(5), 2577–2638.

Ferguson, B.: 2025, Trade frictions in surface water markets, *Technical report*, Working paper.

Fulton, J.: 2021, Almond water footprints: Beyond the one-gallon-per-nut complex, Presentation at the Water Resources & Policy Initiatives (WRPI) Conference (virtual). California State University.
URL: *https://www.calstate.edu/impact-of-the-csu/research/water/Documents/conference/2021/almond-water-footprints-beyond-the-one-gallon-per-nut-complex.pdf*

Fulton, J., Norton, M. and Shilling, F.: 2019, Water-indexed benefits and impacts of California almonds, *Ecological Indicators* .

Gao, Z., Sockin, M. and Xiong, W.: 2020, Economic consequences of housing speculation, *The Review of Financial Studies* **33**(11), 5248–5287.

Geanakoplos, J. and Polemarchakis, H. M.: 1985, Existence, regularity, and constrained suboptimality of competitive allocations when the asset market is incomplete, *Cowles Foundation Discussion Paper 764*, Cowles Foundation, Yale University.
URL: *http://ideas.repec.org/p/cwl/cwldpp/764.html*

Gisser, M. and Sanchez, D. A.: 1980, Competition versus optimal control in groundwater pumping, *Water resources research* **16**(4), 638–642.

Gorback, C. S. and Keys, B. J.: 2025, Global capital and local assets: House prices, quantities, and elasticities, *Technical report*.

Greenwald, B. C. and Stiglitz, J. E.: 1986, Externalities in economies with imperfect information and incomplete markets, *The quarterly journal of economics* **101**(2), 229–264.

Greenwood, R. and Hanson, S. G.: 2015, Waves in ship prices and investment, *The Quarterly Journal of Economics* **130**(1), 55–109.

Hadachek, J., Bruno, E. M., Hagerty, N. and Jessoe, K.: 2026, Externalities of climate adaptation in common-pool groundwater resources, *Journal of Public Economics* **256**, 105602.

Hagerty, N.: 2022, Adaptation to surface water scarcity in irrigated agriculture, *Unpublished, Working Paper* .

Hagerty, N.: 2026, Transaction costs and the gains from trade in water markets, *Review of Economics and Statistics* .

Hasey, J. K., Buchner, R. P., Klonsky, K., Sumner, D., Anderson, N. and Stewart, D.: 2015, Sample costs to establish an orchard and produce english walnuts.

Hayek, F. A.: 1931, *Prices and Production*, George Routledge and Sons, London.

He, Z. and Kondor, P.: 2016, Inefficient investment waves, *Econometrica* **84**(2), 735–780.

Hornbeck, R. and Keskin, P.: 2015, Does agriculture generate local economic spillovers? short-run and long-run evidence from the ogallala aquifer, *American Economic Journal: Economic Policy* **7**(2), 192–213.

Hsieh, C.-T. and Moretti, E.: 2003, Can free entry be inefficient? fixed commissions and social waste in the real estate industry, *Journal of Political Economy* **111**(5), 1076–1122.

Inamda, N.: 2025, India's data centre boom confronts a looming water challenge, *BBC* . Accessed: 2026-02-07.

Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Shamsudduha, M., Taylor, R. G., Fallatah, O. and Kirchner, J. W.: 2024, Rapid groundwater decline and some cases of recovery in aquifers globally, *Nature* **625**(7996), 715–721.

Kaldor, N.: 1934, A classificatory note on the determinateness of equilibrium, *The Review of Economic Studies* **1**(2), 122–136.

Korinek, A. and Simsek, A.: 2016, Liquidity trap and excessive leverage, *American Economic Review* **106**(3), 699–738.

Kydland, F. E. and Prescott, E. C.: 1982, Time to build and aggregate fluctuations, *Econometrica: Journal of the Econometric Society* pp. 1345–1370.

Lanteri, A. and Rampini, A. A.: 2023, Constrained-efficient capital reallocation, *American Economic Review* **113**(2), 354–395.

Lee, D. S., McCrary, J., Moreira, M. J. and Porter, J.: 2022, Valid t-ratio inference for iv, *American Economic Review* **112**(10), 3260–3290.

Lorenzoni, G.: 2008, Inefficient credit booms, *The Review of Economic Studies* **75**(3), 809–833.

Lund, J., Medellín-Azuara, J., Durand, J. and Stone, K.: 2018, Lessons from California's 2012-2016 Drought, *Journal of Water Resources Planning and Management* **144**(10), 04018067.

Mankiw, N. G. and Whinston, M. D.: 1986, Free entry and social inefficiency, *The RAND Journal of Economics* **17**(1), 48–58.

Oster, E.: 2019, Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business & Economic Statistics* **37**(2), 187–204.

Pollack, S.: 1998, An analysis of the feasibility of providing federal multi-peril crop insurance to nut tree growers: Almonds, hazelnuts, pecans, pistachios, walnuts, *Report to the risk management agency*, Economic Research Service, U.S. Department of Agriculture.
URL: *https://legacy.rma.usda.gov/pilots/feasible/PDF/nuttrees.pdf*

Povel, P., Sertsios, G., Kosova, R. and Kumar, P.: 2016, Boom and gloom, *The Journal of Finance* **71**(5), 2287–2332.

Reif, J.: 2010, STRGROUP: Stata module to match strings based on their Levenshtein edit distance, Statistical Software Components, Boston College Department of Economics.

Rognlie, M., Shleifer, A. and Simsek, A.: 2018, Investment hangover and the great recession, *American Economic Journal: Macroeconomics* **10**(2), 113–153.

Sbranti, J.: 2014, Oakdale almond grower shares its vision, *Merced Sun-Star* .
URL: *https://www.mercedsunstar.com/news/business/agriculture/article3296142.html*

Scanlon, B. R., Healy, R. W. and Cook, P. G.: 2002, Choosing appropriate techniques for quantifying groundwater recharge, *Hydrogeology journal* **10**(1), 18–39.

Sherrick, B.: 2020, The relationship between inflation and farmland returns.

Shinkre, T. and Hazlett, C.: 2024, Demystifying and avoiding the ols" weighting problem": Unmodeled heterogeneity and straightforward solutions, *arXiv preprint arXiv:2403.03299* .

Shleifer, A. and Vishny, R. W.: 1988, The efficiency of investment in the presence of aggregate demand spillovers, *Journal of Political Economy* **96**(6), 1221–1231.

Shokri, N.: 2026, Iran's biggest centres of protest are also experiencing extreme pollution and water shortages, *ABC Asia* . Accessed: 2026-02-07.
URL: *https://www.abc.net.au/asia/iran-protests-environmental-water/106294696*

Solon, G., Haider, S. J. and Wooldridge, J. M.: 2015, What are we weighting for?, *Journal of Human resources* **50**(2), 301–316.

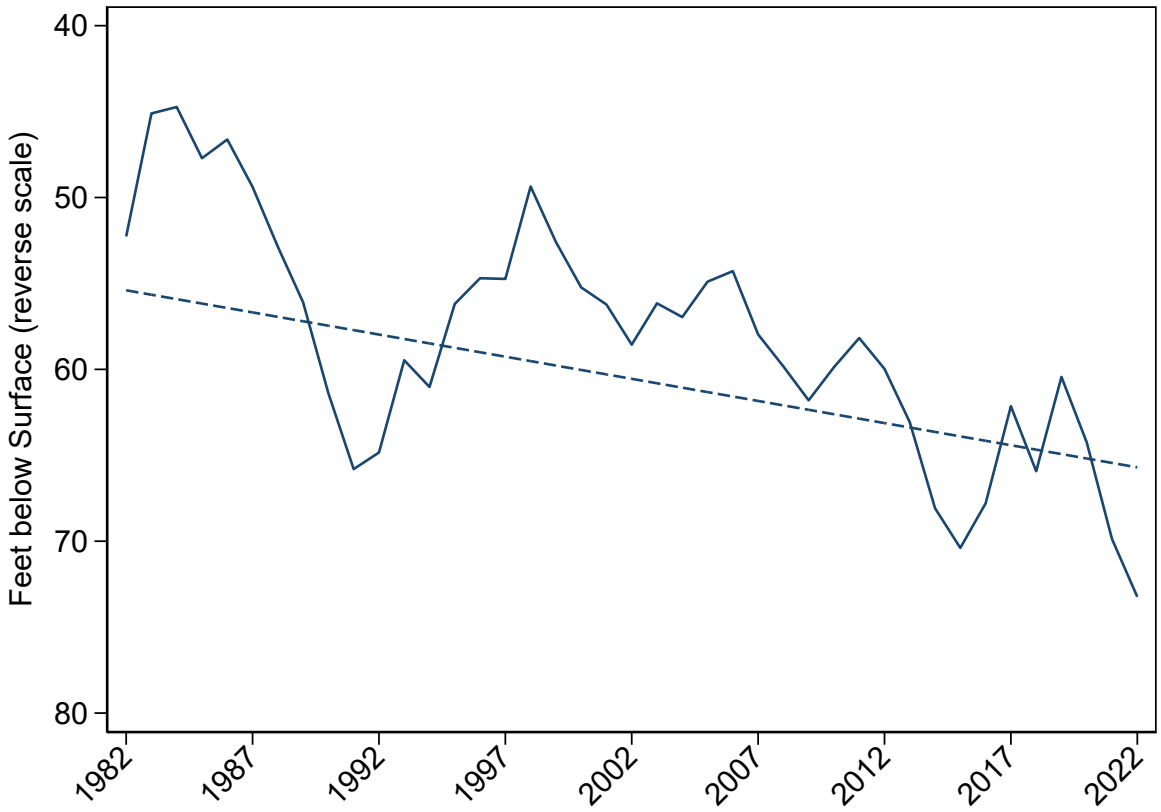
Stock, J. H. and Yogo, M.: 2005, Testing for weak instruments in linear iv regression, *in* D. W. K. Andrews and J. H. Stock (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, Cambridge, pp. 80–108.

U.S. Department of Agriculture, Foreign Agricultural Service: 2025, Production, supply and distribution online (PS&D).

U.S. Environmental Protection Agency: 2025, AP 42, Fifth Edition, Volume I Chapter 9: Food and Agricultural Industries, https://www.epa.gov/air-emissions-factors-and-quantification/ap-42-fifth-edition-volume-i-chapter-9-food-and-0. Last updated May 29, 2025; accessed February 3, 2026.

Yaghmour, M., Haviland, D. R., Fichtner, E. J., Sanden, B. L., Viveros, M., Sumner, D. A., Stewart, D. E. and Gutierrez, C. A.: 2016, Sample costs to establish an orchard and produce almonds.

Figure 1: Aquifer Depth in California.

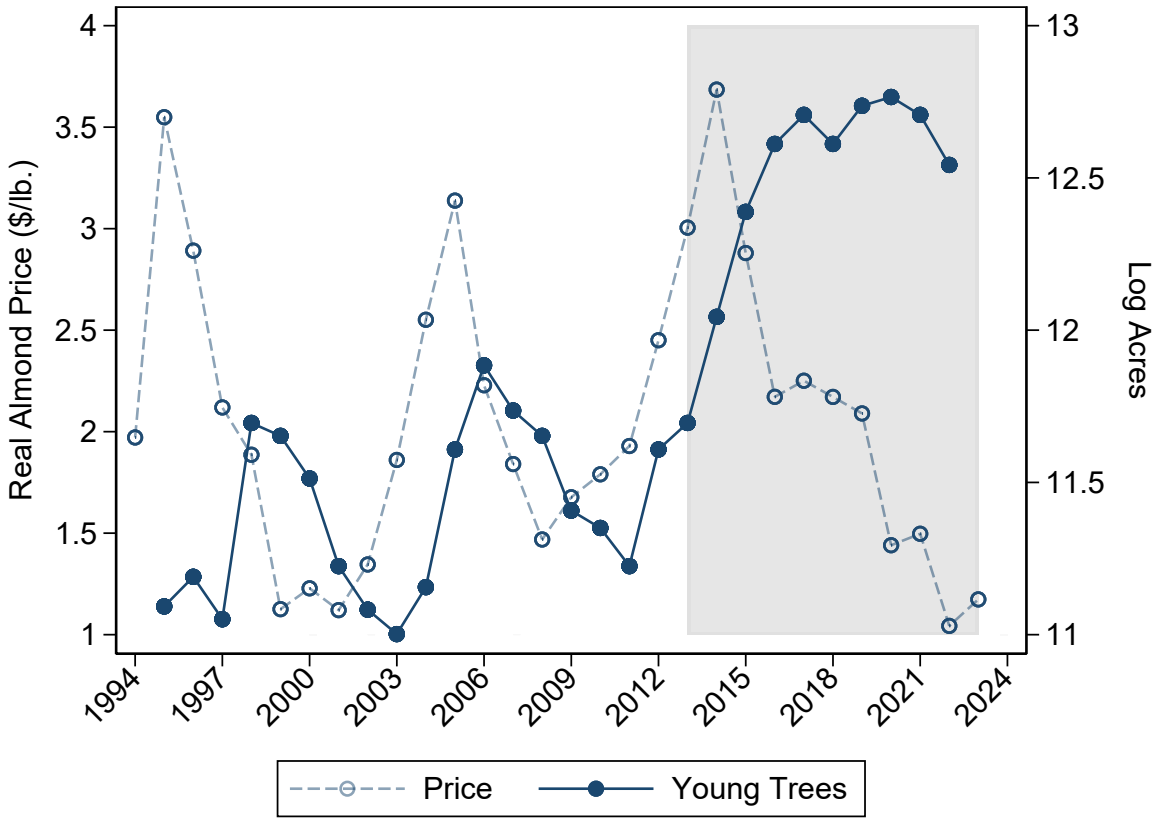


Note: The vertical axis has a reverse scale. The figure presents a time series of the depth to the aquifer across groundwater measurement stations in California. Data are from the California Department of Water Resources (DWR) Periodic Groundwater Level Measurements. To address changes in the composition of wells measured each year, we estimate the pooled regression across wells w and years t

$$AquiferDepth_{w,t} = \alpha_w + \tau_t + \epsilon_{w,t},$$

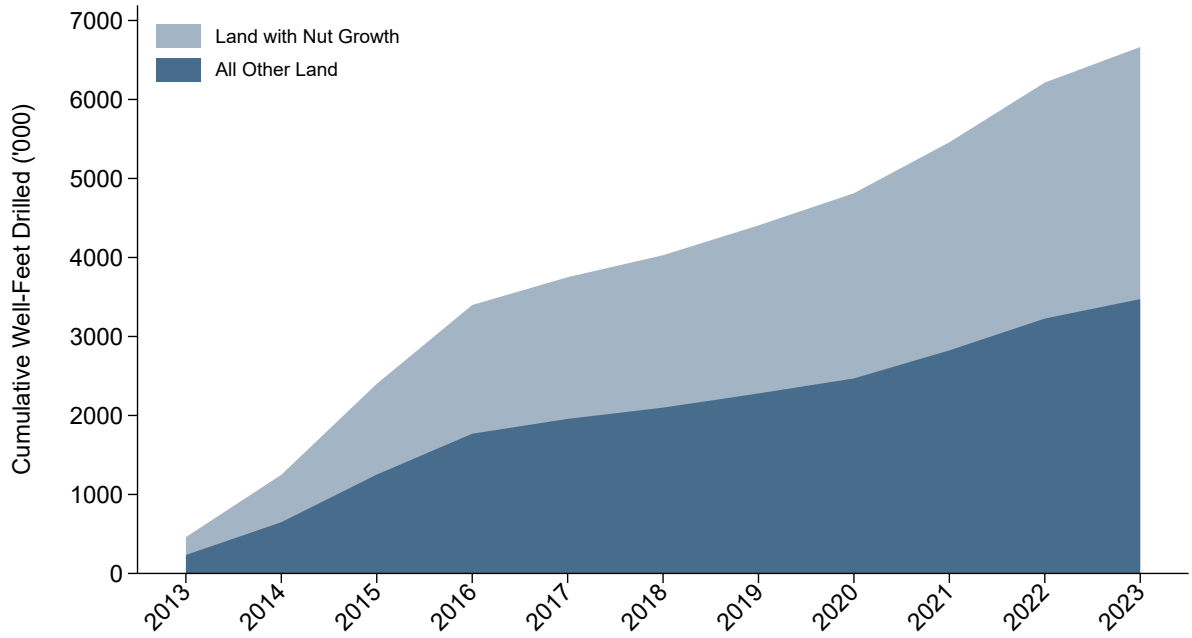
where α_w and τ_t are well and year fixed effects, respectively, and $AquiferDepth_{w,t}$ is the depth to the aquifer, in feet, as measured at w in t. The figure specifically plots the coefficients τ_t, which have the interpretation of a constant-well index. Aquifer depth is expressed in feet below land surface, representing the vertical distance from the land surface to the water table. Increasing depth values correspond to declining groundwater levels and reduced water availability within the aquifer system, which is why the figure has a reverse scale. Section 2 describes our data sets.

Figure 2: Real Nut Price and Investment in New Trees. The Case of Almonds.



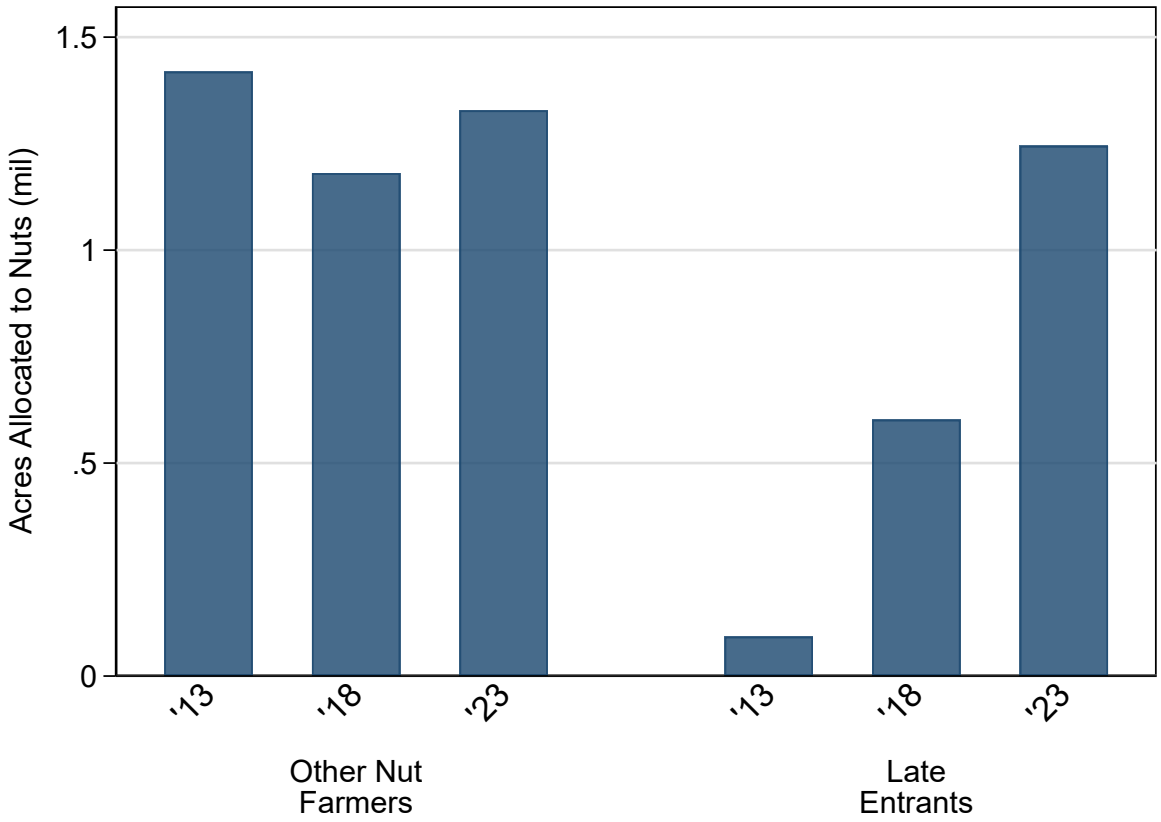
Note: This figure plots real almond prices and acres of young almond trees (i.e., non-bearing acres) in California. Since almond trees take around four years until they can begin to bear fruit, changes in the acres of young almond trees is analogous to investment in almond trees. Gray bars mark our 2013-2023 period of analysis. Almond prices are in real dollars per pound for almonds produced in California, which accounts for around 80% of global almond production. Real prices are computed using the full CPI basket and are in 2010 dollars. Data on almond prices and acres of bearing or non-bearing trees are from the Almond Board of California. Section 2 describes our data sets.

Figure 3: Decomposing Well Drilling on California Agricultural Land by Nut Growth.



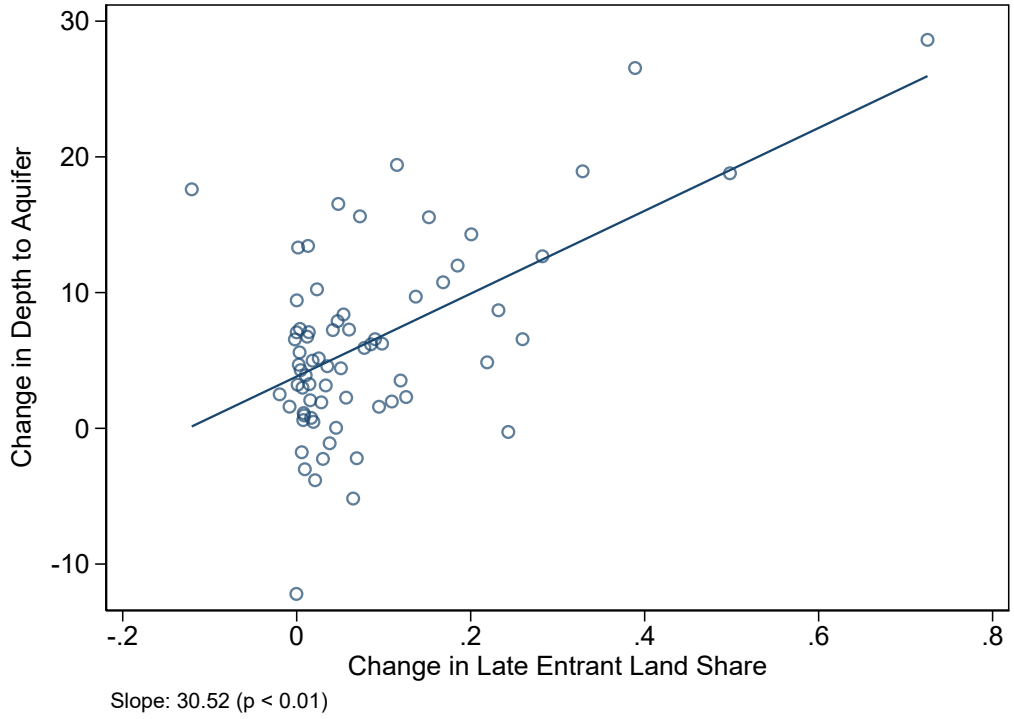
Note: This figure plots the cumulative number of feet of groundwater wells drilled from 2013 through 2023 on land in agricultural California. The figure decomposes this number into: the part from parcels with growth of at least 5 pp in the share of land allocated to nuts over 2013-2023 (Land with Nut Growth); and the part from the remaining parcels in the data set (All Other Land). The set of parcels used to construct the figure are those in the raw CoreLogic-Cropscape data set after imposing the standard filters that (a) the parcel have no less than half its land allocated to active or fallowed cropland in 2023 and (b) that the parcel is not owned by the federal or state government in 2023. Section 2 describes our data sets.

Figure 4: Nut Cropland of Existing Nut Growers and Late Entrants.



Note: The figure plots acres of nut cropland (in millions) in a given year on parcels that were owned by existing nut growers or late entrants as of that year. Per the note to Table 1, late entrants are defined as owners that allocated less than 10% of their land in the CoreLogic-Cropscape data set to nuts in 2013 or did not own cropland that year. The set of parcels used to construct the figure are those in the raw CoreLogic-Cropscape data set after imposing the standard filters that the parcel have no less than half its land allocated to active or fallowed cropland in 2023 and that the parcel is not owned by the federal or state government in 2023. Section 2 describes our data sets.

Figure 5: Late Entrant Nut Expansion and Change in Aquifer Depth.



Note: The figure plots the relation between the change in depth to the aquifer and tree nut expansion by late entrants over 2013-2023. The plot is binned, and each marker corresponds to around 20 groundwater measurement stations. A station consists of roughly 680 acres, on average, and is the most granular geographic unit at which aquifer depths are observed. The variable on the horizontal axis is the 2013-2023 change in the share of land in a station comprised of tree nuts on parcels that late entrants owned in 2018, the midpoint of the boom (i.e., $\Delta LateEntrantLandShare_s$). As in Figure 4, late entrants are defined as allocating less than 10% of their land in the CoreLogic-Cropscape data set to tree nuts in 2013. The variable on the vertical axis is the 2013-2023 change in depth to the aquifer, in feet ($\Delta AquiferDepth_s$). So, higher values imply a larger drop in groundwater levels. Stations are weighted by their land area. The figure excludes a small minority of stations with extreme changes in depth to the aquifer such that they cannot fit on the figure. The small minority of cases with negative changes in late entrants' land share happen when late entrants acquire nut orchards and tear down the trees (i.e., "shorting" the boom). Section 2 describes our data sets.

Table 1: Summary Statistics.

	Mean	Std. Deviation	Median		Mean	Std. Deviation	Median
$\Delta AquiferDepth_s$	6.111	15.374	2.6	$\Delta PerennialShare_s$	0.006	0.101	0
$WellFeetPerAcre_s$	0.628	1.12	0.156	$\Delta RowShare_s$	-0.027	0.155	0
$\Delta LateEntrantLandShare_s$	0.078	0.126	0.031	$\Delta FallowShare_s$	-0.035	0.104	-0.017
$\Delta OtherNutLandShare_s$	0.034	0.102	0	$\Delta NonCroplandShare_s$	-0.055	0.147	-0.015
$AquiferDepth_{s,13}$	53.016	63.655	28.903	$AbnormalRainfall_s$	-0.201	0.09	-0.182
$PerennialShare_{s,13}$	0.119	0.218	0.011	$SoilErosionFactor_s$	0.291	0.068	0.284
$RowShare_{s,13}$	0.401	0.337	0.359	$SoilWaterStorage_s$	22.479	6.102	22.825
$NutShare_{s,13}$	0.201	0.283	0.028	$SoilPH_s$	7.424	0.537	7.455
$FallowShare_{s,13}$	0.091	0.112	0.058	$DroughtIntensity_{c(s)}$	0.988	0.061	1
$NonCroplandShare_{s,13}$	0.188	0.196	0.11	$HighRegulatoryPriority_{b(s)}$	0.634	0.482	1
$DistanceToProcessor_s$	0.192	0.324	0.085				
Number of Stations: 1,336							

Note: This table summarizes variables at the level of the groundwater measurement station, s, which is our main unit of analysis. Each station consists of roughly 680 acres, on average. The station is the most granular geographic unit at which aquifer depths are observed. Stations are constructed by matching parcels in the CoreLogic-Cropscape data set to their nearest measurement station in the DWR data set. The following variables come directly from the CoreLogic-Cropscape data set: $\Delta LateEntrantLandShare_s$ is the change in acres of nut cropland over 2013-2023 on land owned by late entrants in 2018 divided by the total acres in s, where late entrants are defined as allocating less than 10% of their land in the CoreLogic-Cropscape data set to tree nuts in 2013; $\Delta OtherNutLandShare_s$ is defined similarly in terms of share of land in s not owned by late entrants in 2018; $PerennialShare_{s,13}$ through $NonCroplandShare_{s,13}$ are the 2013 shares of land in s allocated to non-nut perennial crops, row crops, tree nuts, fallow cropland, and non-cropland; and $\Delta PerennialShare_s$ through $\Delta NonCroplandShare_s$ are the 2013-2023 changes in the share of land allocated to the indicated land cover. These variables come from the DWR data set: $AquiferDepth_{s,13}$ is the depth to the aquifer in 2013, in feet; $\Delta AquiferDepth_s$ is the change in this depth from 2013 to 2023; $WellFeetPerAcre_s$ is the quantity of net well-feet drilled over 2013-2023 on parcels in s for non-measurement purposes, divided by the total number of acres in s. The following variables are obtained through the 2023 SSURGO data set and are aggregated to the station level: $SoilErosionFactor_s$ is the K(w) factor, which measures how susceptible soil particles are to detachment by water; $SoilWaterStorage_s$ is the volume of water that the soil can store that is available to plants, in centimeters; and soil pH ($SoilPH_s$). Additional variables are: $DistanceToProcessor_s$ is the average distance to the nearest tree nut processor, in hundreds of miles, across parcels in s, based on almond, pistachio, and walnut trade group web pages; $AbnormalRainfall_s$ is the log difference in annual rainfall over 2012-2016 and annual rainfall over 1895-2012, based on the PRISM data set; $WaterRightShare_s$ is the share of parcels in s associated with an entity with long-term surface water rights, based on the Hagerty (2022) file, which draws on data from the Electronic Water Rights Information Management System and various other sources; $DroughtIntensity_{c(s)}$ is the maximum drought intensity over 2013-2018 in the surrounding county, $c(s)$, as measured by the National Drought Mitigation Center's Drought Severity Combined Index. $HighRegulatoryPriority_{b(s)}$ indicates if the surrounding groundwater subbasin, $b(s)$, is designated as medium or high priority in 2014 by the Sustainable Groundwater Management Act, and, thus, subject to an accelerated timeframe for the introduction of groundwater monitoring and extraction limits. Station-level statistics are weighted by the area of the station. Details are in Section 2 and Appendix A.

Table 2: Effect of Nut Investment on Well Drilling.

	$WellFeetPerAcre_s$		
	(1)	(2)	(3)
$\Delta LateEntrantLandShare_s$	1.487***	1.772***	1.816***
	(0.289)	(0.515)	(0.516)
$\Delta OtherNutLandShare_s$	0.989***	0.925	0.733
	(0.320)	(0.558)	(0.509)
Controls			
Crop Shares & Aquifer Depth, 2013		Y	Y
Changes in Crop Shares, 2013-2023		Y	Y
Soil Characteristics		Y	Y
Drought and Regulation		Y	Y
Pre Period Well Drilling			Y
R-squared	0.037	0.155	0.207
Number of Observations	1,336	1,336	1,336

Note: This table estimates the effect of investment in nut cropland over 2013-2023 on well drilling across geographies according to equation (1),

$$Y_s = \beta \Delta LateEntrantLandShare_s + \delta \Delta OtherNutLandShare_s + \gamma X_s + u_s$$

The unit of analysis is the groundwater measurement station, s. A station consists of roughly 680 acres, on average, and is the most granular geographic unit at which aquifer depths are observed. The variable $\Delta LateEntrantLandShare_s$ is the change in acres of nut cropland over 2013-2023 on land owned by late entrants in 2018, the midpoint of the boom, divided by the total acres in s, where late entrants are defined as allocating less than 10% of their land in the CoreLogic-Cropscape data set to tree nuts in 2013. The variable $\Delta OtherNutLandShare_s$ is defined similarly in terms of share of land in s not owned by late entrants in 2018. The outcome $WellFeetPerAcre_s$ is the quantity of well-feet-per-acre drilled over 2013-2023 on parcels in s for non-measurement purposes, divided by the total number of acres in s. The controls in Crop Shares & Aquifer Depth 2013 are: $AquiferDepth_{s,13}$, $RowCropShare_{s,13}$, $PerennialShare_{s,13}$, $NutShare_{s,13}$, and $FallowShare_{s,13}$ from Table 1, such that $NonCroplandShare_{s,13}$ is the residual. The controls in Changes in Crop Shares, 2013-2023 are: the 2013-2023 changes in share of land allocated to the previous land covers, excluding the change in share of land allocated to nuts as it is spanned by the main independent variables. The controls in Soil Characteristics are: $SoilErosionFactor_s$, $SoilWaterStorage_s$, and $SoilPH_s$. The controls in Drought and Regulation are: $DroughtIntensity_{c(s)}$, $HighRegulatoryPriority_{b(s)}$, and $AbnormalRainfall_s$. Column (3) controls for well-feet-per-acre drilled over 2018-2012 as a pre-period. Additional data details are in the note to Table 1. Stations are weighted by their area. Standard errors clustered by subbasin are in parentheses.

Table 3: Effect of Nut Investment on Aquifer Depth.

	$\Delta AquiferDepth_s$				
	(1)	(2)	(3)	(4)	(5)
$\Delta LateEntrantLandShare_s$	26.618***	45.189***	33.028***	41.191***	34.140***
	(5.518)	(5.879)	(6.651)	(9.435)	(6.101)
$\Delta OtherNutLandShare_s$	17.337***	32.182***	20.989***	28.893***	21.850***
	(5.088)	(6.471)	(6.864)	(8.490)	(6.505)
Controls					
Crop Shares & Aquifer Depth, 2013		Y	Y	Y	Y
Changes in Crop Shares, 2013-2023		Y	Y	Y	Y
Soil Characteristics		Y	Y	Y	Y
Drought and Regulation		Y	Y	Y	Y
Pre-Trend in Aquifer Depth			'02-'12	'08-'12	'02-'10
R-squared	0.063	0.240	0.298	0.259	0.310
Number of Observations	1,336	1,336	777	960	773

Note: This table estimates the effect of investment in nut cropland over 2013-2023 on the change in depth to the aquifer across geographies, and so it is analogous to Table 2. The unit of analysis is the groundwater measurement station, s. The outcome $\Delta AquiferDepth_s$ is the 2013-2023 change in depth to the aquifer in s, and higher values imply a larger drop in groundwater levels. Columns (3)-(5) control for the pre-period change in depth to the aquifer over various time intervals shown in the row Pre-Trend in Aquifer Depth. The sample size falls in these columns because there are fewer repeat-well measurements farther back in time. The remaining notes are the same as in Table 2.

Table 4: Where Late Entrants Expand.

	$\Delta LateEntrantLandShare_s$				$\log(PricePerAcre_{i,t})$	
	(1)		(2)		(3)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
$\log(\widehat{PricePerAcre_s})$			-0.118**	(0.054)		
$AquiferDepth_{s,13}$	0.089**	(0.041)			-0.206***	(0.030)
$PerennialShare_{s,13}$	-0.053	(0.060)			0.164***	(0.056)
$RowShare_{s,13}$	-0.000	(0.082)			-0.112	(0.076)
$NutShare_{s,13}$	-0.156**	(0.067)			0.134**	(0.063)
$NonCroplandShare_{s,13}$	0.280***	(0.099)			0.227***	(0.067)
$DistanceToProcessor_s$	-0.439***	(0.117)			0.034	(0.085)
$SoilErosionFactor_s$	0.077**	(0.034)			-0.074***	(0.027)
$SoilWaterStorage_s$	0.017	(0.029)			0.052**	(0.022)
$SoilAlkalinity_s$	-0.060	(0.041)			-0.104***	(0.029)
$SoilAcidity_s$	-0.032	(0.036)			0.102***	(0.028)
$DroughtIntensity_{c(s)}$	-0.300***	(0.102)			-0.317***	(0.087)
$HighRegulatoryPriority_{b(s)}$	0.002	(0.029)			0.124***	(0.020)
$AbnormalRainfall_s$	-0.032	(0.050)			-0.141***	(0.035)
Unit of Analysis	Station		Station		Parcel-Year	
Year FE					Y	
R-squared	0.129		0.005		0.207	
Number of Observations	1,336		1,336		4,534	

Note: All variables are standardized to have zero mean and unit variance, except the independent variable in column (2). This table estimates the relation between nut expansion by late entrants over 2013-2023 and other variables related to the value of agricultural land. The unit of analysis in columns (1)-(2) is the groundwater measurement station, s, and that in column (3) is a pair of parcels, i, and years, t. In column (3), the sample is restricted to parcels in our CoreLogic-Cropscape data set that transact over 2000-2023 in the stations that comprise our main analysis sample, excluding: foreclosures and short sales; land purchased by late entrants that experiences nut growth; and transactions with abnormally low prices below \$3,000 in 2020 dollars. The outcome, $\log(PricePerAcre_{i,t})$, is the log of sale price per acre, and $\log(\widehat{PricePerAcre_s})$ is the area-weighted average of the fitted values from column (3) across all transacted land in station s over 2000-2023. Note that, for consistency with columns (1)-(2), all independent variables in column (3) are defined at the station level. This is why we include a year fixed effect but not a parcel or station level fixed effect. The remaining variables are defined in the note to Table 1, except that $SoilAlkalinity_s$ equals $\max\{SoilPH_s - 7, 0\}$ and $SoilAcidity_s$ equals $\max\{7 - SoilPH_s, 0\}$. The remaining variables are defined in the note to Table 1. Heteroskedasticity robust standard errors are in parentheses.

Table 5: Instrumental Variable Estimates.

	$\Delta LateEntrantLandShare_s$	$WellFeetPerAcre_s$		$\Delta AquiferDepth_s$	
	(1)	(2)	(3)	(4)	(5)
$\Delta LateEntrantLandShare_s$		1.772***	3.606***	45.189***	96.998***
		(0.515)	(1.188)	(5.879)	(21.519)
$DistanceToProcessor_s$	-0.186***				
	(0.059)				
$\Delta OtherNutLandShare_s$	-0.713***	0.925	2.183**	32.182***	67.698***
	(0.040)	(0.558)	(0.978)	(6.471)	(16.560)
Station Controls	Y	Y	Y	Y	Y
Estimator	OLS	OLS	2SLS	OLS	2SLS
First-stage F			33.5		33.5
J-statistic (p-value)			0.866		0.445
Number of Observations	1,336	1,336	1,336	1,336	1,336

Note: This table assesses the robustness of the main effect in Table 3 by using an instrument for $\Delta LateEntrantLandShare_s$ to estimate equation (1) through 2SLS. The unit of analysis is the groundwater measurement station, s. The main instrument is $DistanceToProcessor_s$, which is the distance between a parcel and the nearest tree nut processor in hundreds of miles, averaged across parcels in s. Column (1) reports the first stage associated with this instrument, corresponding to equation (4),

$$\Delta LateEntrantLandShare_s = \zeta DistanceToProcessor_s + \tilde{\delta}\Delta OtherNutLandShare_s + \tilde{\gamma}X_s + u_s,$$

Columns (3) and (5) report the estimates from the second-stage equation,

$$Y_s = \beta\Delta\widehat{LateEntrantLandShare}_s + \delta\Delta OtherNutLandShare_s + \gamma X_s + u_s.$$

To avoid weak instrument bias, the instrument set consists of $DistanceToProcessor_s$ and its interaction with the baseline set of controls in column (2) of Table 2, which we denote by Station Controls. For reference, columns (2) and (4) report the associated OLS estimates. The first-stage F-statistic is Cragg-Donald. The remaining notes are the same as in Table 2.

Table 6: Robustness to Within-Station Analysis.

	$WellFeetPerAcre_{i,13\text{-}23}$		
	(1)	(2)	(3)
$LateEntrant_{o(i)}$	0.145***	0.151***	0.147***
	(0.048)	(0.051)	(0.051)
$WellFeetPerAcre_{i,08\text{-}12}$			0.124***
			(0.016)
Parcel-Level Controls		Y	Y
Station FE	Y	Y	Y
R-squared	0.264	0.265	0.275
Number of Observations	51,200	51,200	51,200

Note: This table assesses robustness to bias from station-level unobserved variables in equation (2). The unit of analysis is the land parcel, i. The sample restricts to parcels with growth in the share of land allocated to nuts over 2013-2023. Specifically, the table estimates the difference in groundwater well drilling intensity over 2013-2023 between late entrants and existing nut growers with a station fixed effect, as in equation (5),

$$WellFeetPerAcre_{i,13\text{-}23} = \beta LateEntrant_{o(i)} + \gamma X_i + \alpha_{s(i)} + u_i,$$

where $LateEntrant_{o(i)}$ indicates if the owner of the parcel in 2018, indexed by $o(i)$, is a late entrant to the nut boom, defined as in Table 2; and $\alpha_{s(i)}$ is a station fixed effect. The variable $WellFeetPerAcre_{i,13\text{-}23}$ is the number of well-feet drilled per acre over 2013-2023 on i for non-measurement purposes. Parcel-level controls are the: 2013 shares of land allocated to row crops, non-nut perennial crops, tree nuts, and fallow cropland (non-cropland is the residual); and the 2013-2023 change in share of land allocated to non-nut perennial crops, row crops, fallow cropland, and non-cropland. Parcels are weighted by their area. Heteroskedasticity robust standard errors are in parentheses.

Table 7: Margins of Adjustment. Surface Water Rights and Water Consumed.

	$WellFeetPerAcre_{i,13\text{-}23}$	$HasSurfaceWater_i$	$\log(TotalWater_{i,13\text{-}23})$
	(1)	(2)	(3)
$LateEntrant_{o(i)}$	0.151***	-0.011**	0.013***
	(0.051)	(0.005)	(0.004)
Parcel-Level Controls	Y	Y	Y
Station FE	Y	Y	Y
R-squared	0.265	0.881	0.808
Number of Observations	51,200	51,200	51,200

Note: This table assesses the margins of adjustment by which existing growers drill less than late entrants within a given station. The general specification is the same as in Table 6, and, as in Table 6, the sample restricts to parcels with growth in the share of land allocated to nuts over 2013-2023. For reference, column (1) repeats the same regression as column (2) of Table 6. The outcome in column (2) is $HasSurfaceWater_i$, which is an indicator for whether i lies within the boundary of an entity with post-1914 rights to surface water (e.g., a water district with rights to a river or reservoir). The outcome in column (3) is the log of $TotalWater_{i,13\text{-}23}$, which is the average annual evapotranspiration on parcel i over 2013-2023, in liters per square meter. Evapotranspiration equals the total amount of water lost through the combination of evaporation and absorption by vegetation (i.e., transpiration), and it is a common measure of consumptive water use as described in the text. Data on surface water rights are from Hagerty (2022) and data on water consumed are from Open ET, with details in Appendix A. The remaining notes are the same as in Table 6.

Table 8: Aggregated Effect of Late Entrants on Aquifer Depth.

	Assumed Replacement Share, λ	
	$\lambda = 1$ (Inelastic Demand)	$\lambda = 0$ (Elastic Demand)
Increase in Aquifer Depth, 2013-2023 (feet)		
Actual	6.11	6.11
Counterfactual without late entrants	5.1	2.6
Share of increase due to late entrants (η)	17%	57%

Note: This table reports the contribution of late entrants to the increase in aquifer depth (i.e., decline in groundwater levels) over 2013-2023, based on the estimates in column (2) of Table 3. The upper panel reports both: the average actual increase in aquifer depth across groundwater measurement stations in the analysis sample, weighting by the area of each station; and the average counterfactual increase without late entrants, again weighting by area. The counterfactual in station s is computed as in equation (6) as

$$CounterfactualDepth_{s,23} = (\Delta AquiferDepth_s - \beta_\lambda \Delta LateEntrantLandShare_s) + AquiferDepth_{s,13},$$

where β_λ is the linear combination of coefficients from column (2) of Table 3 corresponding to the assumed share of late entrants' investment that is replaced by existing owners. The share of the increase in depth to the aquifer over 2013-2023 due to late entrants is then computed as

$$\eta = \frac{ActualGroundwaterLoss_{13\text{-}23} - CounterfactualGroundwaterLoss_{13\text{-}23}}{ActualGroundwaterLoss_{13\text{-}23}},$$

where $CounterfactualGroundwaterLoss_{13\text{-}23}$ and $ActualGroundwaterLoss_{13\text{-}23}$ are aggregates of station-level changes in aquifer depths as in equations (7) and (8). In the $\lambda = 1$ case, β_λ equals the difference in coefficients between $\Delta LateEntrantLandShare_s$ and $\Delta OtherNutLandShare_s$. In the $\lambda = 0$ case, β_λ equals the coefficient on $\Delta LateEntrantLandShare_s$. Details are in Section 7.

Table 9: Assessing the Role of Preference for Short-Term Cash Flows.

	$SellInBust_i$	$RealizedHoldingPeriod_i$	$PotentialHoldingPeriod_i$	$WellFeetPerAcre_{i,\text{13-23}}$
	(1)	(2)	(3)	(4)
$LateEntrantPreBust_{o(i)}$	0.016***	-3.638***	-2.530***	
	(0.005)	(1.028)	(0.236)	
$PotentialHoldingPeriod_i$				-0.018**
				(0.008)
Parcel Controls	Y	Y	Y	Y
Station FE	Y	Y	Y	Y
R-squared	0.411	0.924	0.533	0.296
Number of Observations	51,200	237	11,796	11,796

Note: This table assesses whether late entrants have a shorter holding period. The unit of analysis is the land parcel, i. The regression equation is

$$Y_i = \beta LateEntrantPreBust_{o(i)} + \gamma X_i + \alpha_{s(i)} + u_i,$$

where $LateEntrantPreBust_{o(i)}$ indicates if the owner of the parcel in 2023, the last year before the bust, is a late entrant to the boom, where late entrants are defined as in Table 2; and $\alpha_{s(i)}$ is a station fixed effect. The outcomes in columns (1)-(3) are: $SellInBust_i$, which indicates if i was sold during the bust period of 2024-2025; $RealizedHoldingPeriod_i$, which is defined for parcels that sell in the bust period of 2024-2025 and equals the difference between the year of sale in the bust and the most recent year of sale prior to the bust; and $PotentialHoldingPeriod_i$, which is the holding period if the parcel were to hypothetically sell in 2024. The variables $RealizedHoldingPeriod_i$ and $PotentialHoldingPeriod_i$ are only calculated for parcels with at least 10% of their area allocated to tree nuts in 2013. This restriction avoids cases where the owner converted non-nut cropland to nut orchards over the boom, so that all ownership spells in the sample have a similar interpretation as buy-and-maintain investments as of the start of the boom. Data on sales over 2024-2025 are from Regrid. The remaining notes are the same as in Table 6.

Table 10: Assessing the Role of Ownership of the Common Pool.

	$ShareOfStationOwned_{s,o}$	$WellFeetPerAcre_{s,o}$	$ShareOfBasinOwned_{b,o}$	$WellFeetPerAcre_{b,o}$
	(1)	(2)	(3)	(4)
$LateEntrant_o$	-0.014***		-0.000***	
	(0.003)		(0.000)	
$ShareOfStationOwned_{s,o}$		-0.277***		
		(0.090)		
$ShareOfBasinOwned_{b,o}$				-2.465**
				(1.203)
Station FE	Y	Y		
Subbasin FE			Y	Y
R-squared	0.501	0.143	0.469	0.023
Number of Observations	69,837	69,837	52,711	52,711

Note: This table assesses whether late entrants internalize the aquifer less than existing growers because they own a smaller share of it. Subscripts o, s, and b index owner, station, and subbasin. Each observation is an owner-by-station or owner-by-subbasin pair. The regression equation in columns (1) and (3) is of the form

$$ShareOfGeographyOwned_{o,g} = \beta LateEntrant_o + \alpha_g + u_{o,g}. \tag{14}$$

The outcomes in columns (1) and (3) are the share of the geography g (s for station and b for subbasin) owned by o in 2018, the midpoint of the boom. The independent variable, $LateEntrant_o$, indicates if o is a late entrant. The regression equation in columns (2) and (4) is similar, except that: $ShareOfGeographyOwned_{o,g}$ is now the independent variable; and the dependent variable is now the average of $WellFeetPerAcre_i$ across parcels i owned by owner o in geography g. Standard errors clustered by owner are in parentheses.

Online Appendix

A Data Appendix

A.1 Details on the Cropland Data Layer Data Set

The CDL data are stored as rasters, which can be downloaded from the USDA's website. Each pixel in the image has a color, and the color is associated with a given crop (e.g., pumpkins, radishes) or other agricultural land use (e.g., fallow cropland, open space). We work with a resolution 30m, meaning that each pixel has an area of 900m^2 (0.22 acres).

We begin with the master shapefile of land parcels in California obtained from the Los Angeles County Data Hub, described below. We sort counties by total agricultural acres according to the USDA's 2022 Census of Agriculture, and we restrict attention to the counties within this sorted list that cumulatively account for 95% of agricultural land in California. For each land parcel, we compute the number of pixels associated with a given crop. We group certain types of crops (e.g., vegetables) into a single category, as described below.

To compute the share of land allocated to that crop, we divide by the total number of pixels associated with a parcel. Lastly, we filter the data to parcels of at least one acre, where an acre equals around five pixels given our resolution. Repeating from the text, our "cleaned CDL data set" contains the share of land allocated to a given crop in 2013 and 2023 on all land parcels within the California counties that comprise 95% of the state's agriculture.

We group the raw pixel values in the CDL data into the following categories. The numbers correspond to the values from the USDA's codebook: almonds (75); tree nuts (almonds (75), pistachios (204), walnuts (76), pecans (74)); corn (1, 12, 13); hay (36, 37); soybeans (5); cotton (2); tree fruit (66, 67, 68, 77, 211, 217, 218, 220, 223); rice (3); grassland or pasture (171 (for 2023), 181 (for 2023), 176 (before 2023), 152, 62, 64); forested land (141, 142, 143, 63); open space (121); developed land (122, 123, 124); wetland (190, 195); barren land (131, 65); wheat (22, 23, 24); citrus (212, 72); grapes (69); fallow land (61); and vegetable and non-tree fruit (41, 43, 46, 47,49, 50, 54, 48, 55, 206, 207, 208, 209, 213, 214, 215, 216, 219, 221, 222, 227, 229, 231, 242, 243, 244, 245, 246, 247, 248, 249, 250); other, consisting of the remaining categories. Examples of the "other" category are open water, aquaculture, or perennial ice.

As described in the text, we further group each of these grouped crop types into nut cropland, non-nut perennial cropland, row cropland, fallow cropland, and non-cropland. We define non-nut perennial cropland as the sum of: Tree Fruit, Citrus, Berries, and Grapes. We define row cropland as the sum of: Corn, Hay, Soybeans, Cotton, Rice, and Vegetables. We define non-cropland as the sum of: Grassland, Forested Land, Wetland, Barren Land, Developed Land, Open Land, and Other.

A.2 Details on the CoreLogic Data Set

The CoreLogic Assessment data derive from county tax assessor offices and other third-party sources proprietary to CoreLogic. The tax assessments are annual and contain the full name and mailing address of the land owner, the address and geographic coordinates of the parcels, the number of acres, description of the land use and the assessed value of the property. As described in the text, we merge the tax assessment data from CoreLogic with the Cropland Data Layer data using the unique parcel identifier that is common in both data sets. We call the resulting data set the CoreLogic-Cropscape data set.

We also make use of the CoreLogic Transactions Data Set to evaluate the role of expansion via land already owned versus land purchased. This data set contains all ownership transfers of properties in the U.S. as recorded by the county deeds. We focus on the deeds from January 1st, 2013 to December 31st, 2023. The data contain information such as the name of the buyer, the name of the seller, the date of the transaction, the address of the property, the mailing address of the buyer, and the price at which the property was sold

A.3 Name Grouping Algorithm

We begin with the set of all unique owner full names on all parcels in the merged CoreLogic-Cropscape data set. We group full names of non-individual owners into 23 groups. The majority (14) of these groups are defined by whether the owner's name has common legal terms. For example, all names with "LLC" in them constitute one group, and the remaining ungrouped names with "LP" in them constitute a second group, and the remaining ungrouped names with "Orchards" in them constitute another group, and so forth. The remaining 9 groups are defined by whether the owner's name matches the name of a major investor in California agriculture based on prior knowledge: The Wonderful Company, J.G. Boswell Co, Global Agricultural Properties (a subsidiary of TIAA), E.W. Merritt Farms, Assemi Farms, Asellus, Prudential, Rosedale Farming Group, and Etchegaray Farms. The final step of pre-processing involves harmonizing different spellings of common strings, such as replacing

roman numerals with arabic numbers.

We then apply the Reif (2010) string grouping algorithm to each of the 23 groups separately. The purpose of this step is to simply account for variant spellings of the same owner's name (e.g., "The Wonderful Company" versus "Wonderful Company"). At a high level, the algorithm computes the distance between each pair of owner names, in terms of the minimum number of character changes to transform one name into another.[15] A set of strings is classified as a single owner if their distance is sufficiently low. We take a cautious approach and select a low matching threshold of 0.15 on the unit interval, which, based on manual review of the data, errs on the side of under-grouping.

After accounting for differences in spelling, we next account for the possibility that the same entity may own parcels through various subsidiaries. We use the fact that the mailing address on a parcel's tax assessment often corresponds to the effective owner. An exception to this pattern is the case of owners who delegate tax payments to a property manager, in which case the mailing address is that of the manager, not the owner. We proceed by tabulating the set of all unique mailing addresses associated with greater than five unique owner identifiers, excluding post office boxes. Then, we review each mailing address to verify whether it indeed corresponds to a property investment firm, as distinct from a property manager. If so, then we assign the same owner identifier to each owner name with the same mailing address.

Lastly, we manually inspect the results and correct cases of over or under-grouping that arise during manual inspection. For the largest 20 owners of nut cropland, we also cross-reference the owner's name with information about corporate ownership compiled by two websites, Corporation Wiki and Open Corporates.[16] We use this ownership information to consolidate subsidiaries of the same entity into a single owner identifier.

A.4 Details on the Department of Water Resources Data

As mentioned in the text, data on groundwater well drilling and aquifer depth are from the California Department of Water Resources (DWR) Well Completion Reports and Periodic

[15]Per the documentation, this algorithm: "Calculates the Levenshtein edit distance between all pairwise combinations of strings. The Levenshtein edit distance is defined as the minimum number of insertions, deletions, or substitutions necessary to change one string into the other. For example, the Levenshtein edit distance between 'mitten' and 'fitting' is 3, since the following three edits change one into the other, and it is impossible to do it with fewer than three edits." The algorithm then normalizes the raw distance, and we choose to do so by the shorter of the two strings in question. If the normalized distance is less than a threshold of 0.15 on the unit interval, then the two strings are classified as a match. Full grouping is accomplished by transitivity, such that if string A is matched to string B and string B is matched to string C, then A is matched to C.

[16]The URLs for the websites are: https://www.corporationwiki.com/ and https://opencorporates.com/.

Groundwater Level Measurements, respectively. We access the data on groundwater via the California Natural Resources Agency Open Data Platform.

Well Drilling. We restrict attention to new wells drilled from 2013 through 2023, based on the date at which the work on the well ended. We drop wells for which the planned use is monitoring. We observe the latitude and longitude of the well, which enables us to assign a new well record to a given parcel in our cleaned CDL data. We also observe the well's depth, in feet.

Aquifer Depth. The raw data on aquifer depth derive from measurements obtained from groundwater monitoring wells throughout the state. Depth to the aquifer is defined as the distance from the ground surface to the aquifer's topmost layer. Measurements are not necessarily taken every year. So, we approximate the aquifer depth at a given monitoring well in a given year, t, using the average of any and all measurements taken at that monitoring well within a one-year window around t (i.e., $t-1$, t, and $t+1$).

Grouping Parcels into Measurement Stations. As described in the text, we form geographic units by merging parcels to the nearest groundwater monitoring well, with all parcels matched to that monitoring well comprising a singular geographic unit, which we call the "station". The resulting average station has 680 acres and consists of 12 parcels. In the raw DWR data on aquifer depth described in the previous paragraph, there are 45,488 raw measurement wells, and, thus, a potential sample size of 45,488 stations.

To arrive at our core sample of 1,336 stations, we first restrict the sample of well measurements to those whose quality is deemed "good" by the DWR. This restriction lowers the total number of measurement wells to 42,928. Next, we require that the measurement well match to a land parcel in our main file, which is restricted to parcels in the major agricultural counties of California as described in Appendix A.1. This results in a sample size of 21,455 measurement wells. The sample size falls to 14,420 after imposing our standard filter that the parcel associated with a measurement well: (i) has less than 50% of its area allocated to non-cropland in 2023; and (ii) is not state or federally owned land, which includes tribal land acquired by the federal government. Next, we restrict measurement wells to those for which we observe the change in depth to the aquifer over 2013 through 2023. Since, as described in the previous paragraph, we include all well measurements within a one-year bandwidth of the focal year, this means that we require a well to have measurements at some point in 2012, 2013, or 2014 and also at at some point in 2022, 2023, or 2024. This filter results in a sample of 1,737 measurement wells. The reason for this relatively large drop in sample size is that the DWR appears to phase out wells after a certain number of years of measurement. The final drop in sample size to 1,336 happens because we do not observe all co-variates for all measurement stations in our analysis. The least-well populated variables are those

related to soil.

A.5 Details on Additional data sets

We describe the additional data sets referenced in Section 2.

- **NCREIF.** Data on returns to institutional investment in cropland are from the National Council of Real Estate Fiduciaries (NCREIF). NCREIF is an industry group consisting of large fiduciaries that invest largely on the behalf of tax-advantaged institutions (e.g., pensions). Sherrick (2020) provides a summary of the NCREIF data and its use in studying cropland investing.

- **Drought Severity.** Data on drought severity are from the Drought Severity Combined Index (DSCI) produced by the National Drought Mitigation Center. The index ranges from zero to 500. Higher values imply a more severe drought. The DSCI data are available through the University of Nebraska-Lincoln at the county by week level.

- **Almond Board of California.** Data on aggregate quantity of almond produced, average price per pound of almonds, acres of bearing almond cropland, and acres of non-bearing almond cropland ("young trees") are from the Almond Board of California.

- **Land Parcel Shapefiles.** Data on the master shapefile of land parcels in California are from the Los Angeles County Data Hub. We work with the 2014 shapefile. The data contain the assessor's parcel number (APN) of each parcel. When merging between the cleaned CDL data set, described above, and the CoreLogic data, we work with the APN in CoreLogic associated with the year 2014 or the year closest to 2014 to maximize our match rate between the CDL data and the CoreLogic data. The master shapefile does not contain parcels for the following counties: Colusa, Inyo, Mariposa, Modoc, Plumas, San Luis Obispo, and Siskiyou. Since Colusa lies in the set of counties that collectively comprise 95% of California cropland, we acquire a separate shapefile for Colusa from the county's website to ensure that it is included in our analysis.

- **Groundwater Basin Shapefiles.** Data on the shapefile of groundwater basins and subbasins in California are from the California Department of Water Resources Bulletin 118. We work with the shapefile describing 2003 boundaries. We match each parcel in our data to a subbasin, assigning it to the subbasin with the greatest overlap in cases where the parcel straddles two subbasins. We use the terms "basin" and "subbasin" interchangeably in the text, but our analysis is done in terms of subbasins. There are 515 groundwater basins and subbasins in the raw shapefile, but this number falls after

restricting to subbasins that match to a land parcel in the CDL data set. We obtain the groundwater basin shapefiles from Stanford's EarthWorks library.

- **Soil Characteristics.** We obtain data on soil characteristics from the California Soil Resource Lab (CSRL), which itself sources the data from the USDA-NCSS soil survey data (SSURGO). Per the CSRL's website, the data are available as rasters, which are created by aggregating current SSURGO data within 800m grid cells. The coordinate reference system (CRS) used by these grids is an Albers Equal Area projection centered on the Continental United States, based on the NAD83 datum. The SSURGO data are back-filled with STATSGO where current SSURGO data are not available. We calculate the average of each soil characteristic within each parcel in our main data set. We then aggregate from the parcel level to the station level through an area-weighted average of each soil characteristic across parcels in a station.

- **Rainfall.** We obtain data on annual precipitation at the HUC-12 level from the PRISM data set managed by Oregon State University. We interpret precipitation as rainfall. We separately construct a crosswalk from the parcel level to the HUC-12 level, assigning each parcel to the HUC-12 with which it shares the greatest overlap.

- **Tree Nut Processors.** We obtain data on: almond suppliers that also serve as handlers from the Almond Board of California; pistachio processors that perform hulling and drying from the Administrative Commitee for Pistachios; and walnut handlers that take nuts in-shell from the California Walnut Board. Collectively, we call these entities "tree nut processors". We observe the list and address of processors for almonds and walnuts as of 2026 and for pistachios as of 2014. For processors with only a post office box, we search the web to attempt to find the physical address. Then, for each parcel in the CoreLogic-Cropscape data set, we locate the nearest tree nut processor. We then convert the distance from meters to hundreds of miles. Finally, to avoid reverse causality, we drop all processors of almonds or walnuts established after 2014. Since the processor list for pistachios is already from 2014, we do not need to apply this step to pistachios processors.

- **Water Rights.** Data on surface water rights come from a file produced by Nick Hagerty that draws upon various public sources, such as the State Water Resources Control Board's Electronic Water Rights Information Management System (eWRIMS). See Hagerty (2022) for a detailed description of the data. At a high level, owners of surface water rights have an effectively permanent claim on a particular surface water source, such as a lake, river, stream, or reservoir. In this context, "surface water" excludes

claims to major water projects such as the Central Valley Project (CVP), the State Water Project (SWP), or the Lower Colorado Project. Hagerty (2022) collects data from eWRIMS and a variety of other sources to calculate diversions over 2010-2014 from a particular surface water body made by a particular owner of rights to that body. Owners are separated into small ones whose geographic location is described by a particular point versus larger ones whose location is described by a polygon constructed by Hagerty (2022). We work with the set of polygon owners, which often correspond to water districts. Then, we assign each land parcel in our baseline CoreLogic-Cropscape data set to the polygon with the largest overlap. Finally, we define $HasSurfaceWater_i$, as an indicator for whether i lies within the boundary of an polygon with post-1914 rights to surface water. Practically, these polygons include entities such as water districts with rights to a river or reservoir.

- **Evapotranspiration as Water Consumed**. We follow the convention in the hydrology literature and use evapotranspiration to measure consumptive water use on a parcel, where consumptive use means "water from any source to meet a crop's water consumption" (Boyce et al. 2020). Evapotranspiration equals sum of water that leaves the earth through evaporation plus through transpiration, which is water absorbed by plants that later is released as vapor. We obtain data on evapotranspiration from Open ET, which, per its website, runs six different satellite-driven ET models (ALEXI/DisALEXI, eeMETRIC, geeSEBAL, PT-JPL, SIMS, and SSEBop) to measure evapotranspiration. All models are at 30-meter resolution and use Landsat data. Open ET reports the arithmetic mean across model measures after removing outliers using the Median Absolute Deviation (MAD) method. We obtain the data from Google's Earth Engine. Specifically, for each parcel in our main data set and year over 2013 through 2023, we calculate the total amount of evapotranspiration on that parcel in that year, in liters per square meter. We then take the within-parcel average to obtain the average yearly evapotranspiration. Lastly, we repeat the remark from the text that evapotranspiration is generally not equal to water applied, since some water applied is also lost through runoff and percolation into groundwater.

- **Transactions over 2024-2025**. Since our CoreLogic data end in 2023, we acquire data on transactions occurring over 2024-2025 from Regrid. The data are similar in form and content to the CoreLogic assessment data set, as Regrid also obtains the underlying data from county assessor's and recorder's offices. The data only contains information about the most recent sale of a parcel, not the history of sales. However, since we obtain the Regrid data in April 2026, which covers sales through at most 2025, we can

accurately construct an indicator for whether a parcel is sold over 2024-2025, as used in the paper.

- **Alternative Land Cover Data**. We obtain alternative data on land covers from LandIQ, a private firm that cooperates with the California Department of Water Resources (DWR) to provide the statewide crop mapping. We collect the data in April 2026. At that time, the data cover the calendar years of 2014 and 2016 and the hydrologic years 2018, 2019, 2020, 2021, and 2022. Since we do not observe 2013 or 2023 data, as we do in our main USDA CropScape data set, we instead use the closest year available (i.e., 2014 and 2022). The LandIQ data categorizes nearly 15.4 million acres of land use into more than 50 crop and land use types. Like the CropScape data set, the LandIQ land cover data are based on satellite imagery. Unlike the USDA data, however, the LandIQ data are trained using proprietary data on actual land covers collected by state commodity boards and trade groups, like the Almond Board of California, the California Walnut Board, the California Pistachio Research Board, and the California Dried Plum Board.

  For each year's data, LandIQ and the DWR complete a comprehensive accuracy assessment. Per the description of this process on LandIQ's website, the accuracy assessment first involves setting aside independent ground truthing samples. A stratified random sampling method is used for accuracy assessment sample selection. The data sets are stratified by land cover type and county boundary. The sites set aside for accuracy assessment sites are not used to train the classification algorithm and therefore represent unbiased reference information. The share of correctly-classified almond, pistachio, and walnut polygons in the reference sample in 2014 is 98%, 97%, and 97%, respectively. The corresponding accuracy rates for the 2022 reference sample are 99%, 100%, and 99% for almonds, pistachios, and walnuts, respectively. The accuracy rates in the USDA Cropscape data for almonds, pistachios, and walnuts in 2014 are 88.5%, 72.1%, and 83.4%, respectively, and in 2022 are 91.9%, 89.6%, and 89.1%, respectively.

A.6 Additional Data Cleaning

- We filter the data to exclude parcels with less than half of their area allocated to active or fallowed cropland as of 2023 and to exclude parcels owned by the federal or state government as of 2023.

- The coordinate system used to calculate distances is EPSG:3310 (NAD83 / California Albers), which maximizes the accuracy of distance calculations and reduces error from

the projection of the three-dimensional earth onto a two-dimensional plane.

- We winsorize variables that have a potentially unbounded support at the 2.5% level within each tail. This winsorization rule applies to most variables summarized in Table 1 and Appendix Table A.2 except indicator variables or variables that are defined on a bounded scale, such as soil pH. We also winsorize log transaction prices studied in Table 4 using the same rule.

- As mentioned in the text, we restrict the sample to parcels with growth in the share of land allocated to nuts over 2013-2023 in tables that conduct parcel-level analyses (i.e., Tables 6, 7, 9, and A.12). In this restriction, we drop parcels with growth rates greater than zero but less than or equal to 1 pp, which are almost surely due to measurement error.

B Additional Analysis

B.1 Measurement Error in Late Entrant Share

First, we analytically substantiate the remarks from Section 6.4 about measurement error in late entrants' expansion due to over or under-grouping in the name grouping algorithm. Specifically, we show that such error could affect the estimates through the following channels: attenuation bias, which shrinks the estimated coefficients towards zero; and bias from covariance between under-grouping and the increase in depth to the aquifer, which moves the estimated β away from zero if the covariance is positive but closer to zero if the correlation is negative.

To be precise, we introduce the following notation

$$Y_s \equiv \Delta AquiferDepth_s, \tag{A1}$$

$$M_s \equiv [\Delta LateEntrantLandShare_s, \Delta OtherNutLandShare_s]. \tag{A2}$$

Suppose that M_s is measured with error because of the name grouping algorithm. So, we only observe

$$\widehat{M_s} = M_s + \underbrace{[\xi_s, -\xi_s]}_{\equiv \Xi_s}. \tag{A3}$$

Note the symmetry of ξ around zero in equation (A3). This symmetry obtains from the fact that any change in nut cropland in s must come from a late entrant or an existing grower, by definition. So, if under-grouping in the name grouping algorithm leads us to over-estimate late entrants' expansion by ξ_s, then the same error leads us to under-estimate existing growers' expansion by ξ_s. The sign is reversed in the case of over-grouping.

The OLS estimates of the parameters $[\beta, \delta]$ obtain from

$$\begin{aligned}
[\widehat{\beta}, \widehat{\delta}]' &= [\widehat{M_s}' \widehat{M_s}]^{-1} \widehat{M_s}' Y_s \\
&= [M_s' M_s + M_s' \Xi_s + \Xi_s' M_s + \Xi_s' \Xi_s]^{-1} [M_s + \Xi_s]' Y_s \\
&= \Lambda [\beta, \delta]' + \Lambda [M_s{}' M_s]^{-1} \Xi_s{}' Y_s.
\end{aligned} \tag{A4}$$

where

$$\Lambda = [M_s'M_s + M_s'\Xi_s + \Xi_s'M_s + \Xi_s'\Xi_s]^{-1}[M_s{}'M_s] \tag{A5}$$

measures the degree of attenuation bias. The second term in equation (A4) captures bias from the covariance between error from under-grouping and the change in depth to the aquifer, $\Xi_s{}'Y_s$. Since $\Xi_s = [\xi_s, -\xi_s]$, the bias in β will be biased away from zero if ϵ_s and Y_s have positive covariance, while they will be biased towards zero with negative covariance.

So, under-grouping in the name grouping algorithm would explain our results only if the propensity to under-group owner names from the CoreLogic Assessment data set is positively related to the change in depth to the aquifer over 2013-2023 from the California DWR data set. An economic explanation for such a relation is unlikely, and so any such relation must be purely spurious. However, the robustness of the main results across different geographic units and data sets also makes spurious correlation an unlikely explanation.

Nevertheless, we more-directly evaluate the sensitivity of our results to measurement error by restricting our analysis to owners whose names are plausibly less prone to under-grouping. In particular, while we go to great lengths to try and avoid such cases of under-grouping for institutional owners, we cannot do the same for individuals simply because there are no third-party resources to verify whether two variant spellings of the same person's name indeed correspond to the same person. Since individual owners plausibly own smaller properties, we can assess robustness to this form of misclassification by restricting our sample to land parcels larger than some threshold, which we set at 25 or 50 acres.[17]

The resulting estimates in Appendix Table A.4 are very similar to the baseline estimates in Tables 2 and 3. Appendix Figure A.7 further verifies that the contribution of late entrants to aggregate tree nut expansion, shown in Figure 4, is also robust to dropping small parcels. Overall, the similarity obtained when dropping small parcels supports the main results and suggests that they are not driven by over-classification of individual owners as late entrants.

[17]We can use the corporate buyer indicator in the CoreLogic Transactions data set to substantiate this approach. Accordingly, 46% (51%) of transactions over 2000-2023 on parcels at least 25 (50) acres in size had a corporate buyer, versus 25% (27%) for smaller parcels.

B.2 Sensitivity to Measurement Stations as Unit of Analysis

We conduct our analysis of aquifer depth at the level of the groundwater measurement station to increase sample size, and, thus, power, as we described in Section 2. To address concerns about this geographic definition, we replicate a version of Table 3 at the level of the detailed-analysis-unit-by-county (DAUCO). A DAUCO is a geography defined by pairs of "detailed analysis units" as defined by the California DWR and counties, and it has been studied as the unit of analysis in other papers about water in California. Hadachek et al. (2026), for example, perform a DAUCO-level analysis to study the effect of weather shocks on groundwater depletion. Moreover, Ferguson (2025) performs a DAUCO-level analysis to study frictions to trade in surface water markets.

After restricting to DAUCOs for which we observe both expansion by late entrants to the nut market and repeat well measurements in both 2013 and 2023, our sample size falls to just 108. So, we simply examine a DAUCO-level scatterplot of the relation between the 2013-2023 changes in depth to the aquifer vs. share of land owned by late entrants. The result in panel (a) of Appendix Figure A.8 shows a positive and significant association between change in aquifer depth and expansion by late entrants across DAUCOs. This result continues to hold after residualizing both variables by DAUCO-level controls, shown in panel (b). While we hesitate to interpret the magnitude given the small sample size, we note that the best fit line has a slope on par with the estimated 2SLS coefficient in Table 5.

The fact that we obtain similar results at a different unit of analysis suggests that our main finding is not driven by spurious correlation or by how we define measurement stations, supporting its validity.

B.3 Errors in Satellite Land Cover Classifications.

Aware of the possibility that our baseline USDA CropScape data set may mis-classify specific crop types, we replicate our results using data from a private vendor, LandIQ. Like CropScape, the LandIQ data are based on satellite images. Unlike CropScape, however, the LandIQ data are trained using proprietary data from commodity boards or trade groups, such as the Almond Board of California. Consequently, LandIQ reports accuracy rates above 97% for specific crop types in its validation tests, as we describe in Appendix A.

The results in Appendix Table A.5 are similar to our main findings in column (2) of Tables 2 and 3. In particular, reducing nut late entrants' nut expansion by 10% of a station's area and replacing it with expansion by existing growers corresponds to 0.09 more well-feet per acre and a 2.01 foot greater increase in depth to the aquifer. By comparison, Tables 2 and 3 imply that this same exchange corresponds to a fairly-similar increase of 0.08 well-feet per acre and a 1.3 foot greater increase in aquifer depth. As discussed in Section 7, we obtain similar implications for aggregate groundwater levels based on the LandIQ data as with our baseline CropScape data set. Appendix Figure A.9 shows how the aggregate dynamics documented in Figures 3 and 4 are similar when calculated using the LandIQ data, although we note that the LandIQ data imply a smaller contribution of land with nut growth to overall well drilling (35%) than do the USDA CropScape data (48%).

This similarly strongly supports the validity of the main results and provides evidence that they are not driven by measurement error in the USDA's crop classifications.

B.4 Assessing the Role of Water Applied

We stated in the text that the finding from Table 7 that late entrants consume more water is unlikely to fully explain why their expansion corresponds to a greater decline in groundwater levels. We reach this conclusion from the following back-of-envelope calculation. First, we aggregate the estimated difference of 5.2 liters per square meter per year from Appendix Table A.12 across the 2013-2023 period through multiplication by 11. We then convert from liters per square meter to feet, which implies that replacing one late entrant with an existing grower would reduce water consumption by 0.2 feet. Since the average station has 12 parcels, this implies that performing this replacement through the average station would reduce water consumption by 2.4 feet. By contrast, comparing the coefficients on the change in late entrant vs. existing grower land shares in Table 3 implies that exchanging the two types of nut growers would reduce the drop in depth to the aquifer by 13 feet. However, we note that this is a very rough calculation and it is still possible that differences in total water applied can account for the majority of the difference in the effects of late entrants versus existing growers in Table 3.

B.5 Weights of Heterogeneous Treatment Effects

Since at least Angrist (1998), it has been known that adding controls to a regression such as that in equation (1) can change the weights across treatment effects that comprise the coefficient of interest, β. Thus, the fact that the controlled specifications in Tables 2 and Table 3 yield somewhat larger estimates than the uncontrolled specifications may not actually mean that the controls reduce omitted variables bias, but, rather, that adding controls up-weights the contribution of stations with a larger treatment effect. We allow for this possibility by re-estimating equation (1) after additionally including the interactions between $\Delta LateEntrantLandShare_s$ and all of the other controls in the regression, after first re-centering them to have means of zero. This procedure is one of several procedures commonly used to address the potential for changing treatment weights, as recently summarized by Shinkre and Hazlett (2024). The resulting estimates in Appendix Table A.6 are quite similar to their counterparts in Tables 2 and Table 3, reinforcing our earlier interpretation that including controls reduces omitted variables bias.

B.6 Expectations of Future Nut Prices

We provide two pieces of evidence consistent with orchard owners "over-estimating" the persistence of tree nut prices at the beginning of the boom, where we use the term "over-estimating" in the ex-post sense and do not wish to take a stance on the specific forecasting friction that leads firms to over-estimate.

Price Persistence in Cost Manuals

The first piece of evidence comes from publicly-available owner cost manuals from the University of California Agriculture and Natural Resources Cooperative Extension. These manuals provide highly detailed guides about the costs of establishing an almond, pistachio, or walnut orchard in a given year, with the intent of helping orchard owners make well-informed investment decisions. Importantly, these manuals provide estimates of the net present value of establishing an orchard under various scenarios for the future path of tree nut prices.[18]

The first notable feature of these estimates is that they are based on a fixed nut price over the lifespan of the orchard, which implicitly assumes a high degree of persistence that is not found in the data.

The second notable feature is that the future nut price required for growers to break even must lie close to the current price, again, even in boom years. Potentially, this framing may shift growers' priors towards expecting the same price in the future.

The fact that orchard owners − especially late entrants − expand during the boom despite these publicly available estimates of break-even prices suggests that they hold relatively optimistic expectations about the path of future nut prices.

Nut Income Implied by Capitalization Rates

The next piece of evidence comes from expectations of future returns to nut investing implied by the value of nut cropland. We use the NCREIF data set for this exercise, which, per Section 2, contains information about the returns on the almond and pistachio portfolios of fiduciaries that invest on behalf of tax-advantaged institutions. Valuations are based on appraisals by professionals affiliated with the American Society of Farm Managers and Rural Appraisers and are intended to represent market value.

In the forecasting regression, we pool both nut types and include fixed effects for nut type to absorb differences in the stationary component of expected returns across nut types.

[18]Example manuals for the case of almonds in 2016, pistachios in 2015, and walnuts in 2015 appear in Yaghmour et al. (2016), Brar et al. (2015), and Hasey et al. (2015), respectively. The complete collection of manuals can be found at: `https://coststudies.ucdavis.edu/archived/commodities`.

We work at the annual frequency because of strong seasonality in the raw, quarterly data, as described earlier in Appendix A. We aggregate by taking the average quarter-on-quarter arithmetic return across quarters in year t, multiplying by four. We then convert the aggregated arithmetic returns to geometric returns.

Table A.10 reports the regression results. Using some valuation metric for nut type n in year t, we forecast total and income returns on the nut cropland portfolio for type n from t through $t + 3$. Higher valuations of almond and pistachio cropland predict lower total returns, shown in column (1). Low capitalization rates (i.e., net operating income divided by property value) also predict lower returns, consistent with present-value logic applied to dividend-price ratios in the stock market (Campbell and Shiller 1989). What is more challenging to rationalize with present-value logic is that higher property valuations and lower capitalization rates also predict lower *income* return, shown in columns (3)-(4). Greenwood and Hanson (2015) find a similar pattern using present-value regressions in the context of investment in dry bulk shipping. Note that income equals revenue from nut sales minus operating expenses.

This finding suggests that market participants − at least those in the portfolios of fiduciaries that invest on tax-advantaged institutions − over-estimate the persistence of nut prices during the boom. Consistent with this view, Appendix Figure A.4 shows that the valuation of nut cropland in the NCREIF portfolio tends to rise sharply and to remain elevated during nut booms over the last three decades.

C Mathematical Appendix

Proof of Remark 1

Rewrite problem (11) recursively as

$$V_1(D_{i,1}) = \max_{w_{i,1}} \left\{ p_1 \mu A(w_{i,1}) K - w_{i,1} c(D_{i,1}) + e^{-r} \mathbb{E}\left[V_2(D_{i,2}) \right] \right\}, \tag{A6}$$

where

$$V_2(D_{i,2}) = \max_{w_{i,2}} \left\{ p_2 A(w_{i,2}) K - w_{i,2} c(D_{i,2}) \right\}, \tag{A7}$$

and $D_{i,t}$ evolves as in equation (10),

$$D_{i,2} = D_{i,1} + w_{i,1} + \theta \left[\phi w_{i,1} + \int_{\phi}^{1} w_{j \neq i,1} dj \right] - z_{i,1}.$$

We introduce the scalar $\mu > 0$ to analyze a multiplicative increase in $A(\cdot)$. For convenience, we have also set $K = 1$.

Let $w_{i,2}^*$ denote the solution to $V_2(D_{i,2})$ and let F denote the first-order condition,

$$F = p_2 A'(w_{i,2}^*) - c(D_{i,2}) \equiv 0. \tag{A8}$$

Note that

$$\partial F / \partial p_2 = A'(w_{i,2}) > 0, \tag{A9}$$

$$\partial F / \partial w_{i,1} = -c'(D_{i,2})(1 + \theta\phi) < 0, \tag{A10}$$

$$\partial F / \partial \phi = -c'(D_{i,2}) \theta (w_{i,1} - w_{j,1}). \tag{A11}$$

Since Remark 1 hypothesizes that ϕ is small, we analyze equation (A11) for the case $\phi \to 0$. In this limiting case, symmetry implies $w_{i,1} = w_{j,1}$ for all $i \neq j$, since all growers are then nearly infinitesimal. So, equations (A9)-(A11) respectively imply that

$$\partial w_{i,2}^* / \partial p_2 > 0, \;\; \partial w_{i,2}^* / \partial w_{i,1} < 0, \;\; \partial w_{i,2}^* / \partial \phi = 0, \tag{A12}$$

which follow from the implicit function theorem and the concavity of F.

Next, let $w^*_{i,1}$ denote the solution to $V_1(D_{i,1})$ and let G denote the first-order condition,

$$G = p_1 \mu A'(w^*_{i,1}) - c(D_{i,1}) - e^{-r}\mathbb{E}[w^*_{i,2} c'(D_{i,2})(1+\theta\phi)] \equiv 0. \tag{A13}$$

Note that

$$\partial G/\partial r = e^{-r}\mathbb{E}[w^*_{i,2} c'(D_{i,2})(1+\theta\phi)], \tag{A14}$$

$$\partial G/\partial \mu = p_1 A'(w^*_{i,1}), \tag{A15}$$

$$\partial G/\partial p_2 = -e^{-r}\mathbb{E}[(\partial w^*_{i,2}/\partial p_2) c'(D_{i,2})(1+\theta\phi)], \tag{A16}$$

$$\begin{aligned}\partial G/\partial \phi = -e^{-r}\mathbb{E}[&w^*_{i,2} c''(D_{i,2})(1+\theta\phi)\theta[w^*_{i,1} - w_{j,1}] + \\ &w^*_{i,2} c'(D_{i,2})\theta + (\partial w^*_{i,2}/\partial\phi) c'(D_{i,2})(1+\theta\phi)].\end{aligned} \tag{A17}$$

Again taking the limiting case $\phi \to 0$ and using inequality (A12), equation (A17) simplifies to

$$\partial G/\partial \phi = -e^{-r}\mathbb{E}[w^*_{i,2} c'(D_{i,2})\theta]. \tag{A18}$$

Equations (A14) and (A15) imply that

$$\partial w^*_{i,1}/\partial r > 0, \quad \partial w^*_{i,1}/\partial \mu > 0, \tag{A19}$$

Likewise, equation (A16) implies that

$$\partial w^*_{i,1}/\partial p_2 < 0, \tag{A20}$$

after substituting from inequality (A12). Lastly, equation (A18) implies that

$$\partial w^*_{i,1}/\partial \phi < 0. \tag{A21}$$

Together, inequalities (A19), (A20), and (A21) confirm the statements made in the remark: $w^*_{i,1}$ is increasing in r, decreasing in ϕ when ϕ is small, decreasing in p_2, and increasing in a multiplicative increase in $A(\cdot)$ at time 1. This completes what needed to be shown.

Relation between Shorter Holding Periods and Higher Discount Rates

In Section 8.1, we noted that having a shorter holding period incentivizes extraction at time 1 if doing so reduces property prices by less than it reduces the present discounted value of nut profits. To make this point explicit, we modify problem (11) such that growers exit the market at the beginning of time 2 with exogenous probability ψ. If growers exit, then they sell their property to an unmodeled investor who has elastic demand and is willing to pay $P(D)$, where $P(\cdot)$ is decreasing. This investor's willingness-to-pay may not necessarily be equivalent to the time-2 cash flows from the nut orchard if, for example, the investor converts the nut orchard to a less water-intensive land use. The grower's problem becomes

$$V_1(D_{i,1}) = \max_{w_{i,1}} \left\{ p_1 A(w_{i,1})K - w_{i,1}c(D_{i,1}) + e^{-r}\mathbb{E}\left[V_2(D_{i,2})\right] \right\}, \tag{A22}$$

$$V_2(D_{i,2}) = \begin{cases} P(D_{i,2}), & w.p. \quad \psi \\ \max_{w_{i,2}} \left\{ p_2 A(w_{i,2})K - w_{i,2}c(D_{i,2}) \right\} & w.p. \quad 1-\psi \end{cases} \tag{A23}$$

and $D_{i,t}$ evolves as in equation (10). Let $w^*_{i,1}$ denote the solution to $V_1(D_{i,1})$ and let G denote the associated first-order condition,

$$G = p_1 \mu A'(w^*_{i,1}) - c(D_{i,1}) - e^{-r}(1+\theta\phi)\left(\psi P'(D_{i,2}) + (1-\psi)\mathbb{E}[w^*_{i,2}c'(D_{i,2})]\right) \equiv 0. \tag{A24}$$

Note that

$$\partial G/\partial\psi = -e^{-r}(1+\theta\phi)\left(P'(D_{i,2}) - \mathbb{E}[w^*_{i,2}c'(D_{i,2})]\right). \tag{A25}$$

If

$$P'(D_{i,2}) < \mathbb{E}[w^*_{i,2}c'(D_{i,2})], \tag{A26}$$

then $\partial G/\partial\psi > 0$, which, by concavity of G and the implicit function theorem, implies that

$$\partial w^*_{i,1}/\partial\psi > 0. \tag{A27}$$

Condition (A26) is roughly equivalent to saying that extraction at time 1 reduces property prices by less than it reduces the present discounted value of nut profits. If this condition holds, then having a shorter holding period incentivizes extraction, which is equivalent to the result (A27). This substantiates the statement made in the text.

Endogenizing K in the Framework

We sketch the time-0 problem of planting the orchard, K, subject to a convex planting cost $\Phi(K)$. The value function is

$$V_0(D_0) = \max_K \left\{ -\Phi(K) + e^{-r}\mathbb{E}\left[V_1(D_{i,1}, K)\right] \right\}, \tag{A28}$$

where V_1 is defined in the proof of Remark 1. The first-order condition is

$$H = -\Phi'(K) + \sum_{t=1}^{t=2} e^{-rt}\mathbb{E}\left[p_t A(w_{i,t})\right] \equiv 0. \tag{A29}$$

Let K^* denote the solution. Differentiating equation (A29) with respect to p_2 and applying the implicit function theorem yields

$$\partial K^*/\partial p_2 \propto \mathbb{E}\left[A(w^*_{i,2}) + p_2 A'(w^*_{i,2})(\partial w^*_{i,2}/\partial p_2) + e^r p_1 A'(w^*_{i,1})(\partial w^*_{i,1}/\partial p_2)\right], \tag{A30}$$

where $\propto$ denotes proportionality.

The term $A(w^*_{i,2})$ on the right side of equation (A30) captures the direct incentive of persistently-high nut prices to plant orchards. The middle term, $p_2 A'(w^*_{i,2})(\partial w^*_{i,2}/\partial p_2)$, is the additional incentive from the fact that growers will consequently enhance their yield at time 2 by applying more groundwater. The third term, $e^r p_1 A'(w^*_{i,1})(\partial w^*_{i,1}/\partial p_2)$, is attenuation from the fact that, to preserve the aquifer for time 2, growers under-irrigate their orchards at time 1.

Recapping, equation (A30) supports the statement from the text that having more optimistic expectations could explain why late entrants expand more in the first place.

D Additional Institutional Details

We elaborate on the institutional details described in Section 3. The last heading discusses tree nut futures.

California Water Sources

California's dry climate necessitates that cropland be irrigated, including and especially tree nut cropland. Broadly, there are four methods that growers could use to acquire new water sources for irrigation.

- The first method is to drill groundwater wells to tap water stored in aquifers, as described in the text. The remaining four methods involve acquiring surface water. Since our paper concerns groundwater, we provide only a high-level overview here.
- The second method is to contract with one of the major water projects in California, such as the State Water Project, the Central Valley Project, or the Lower Colorado Project. Water from these projects is highly regulated and subject to quantity restrictions in time of drought.
- The third method is to apply for rights to surface water stored in rivers, reservoirs or other natural bodies of water, or, similarly, to buy land that already has such rights. As noted by Hagerty (2026), such surface water rights, while technically regulated by California's Department of Water Resources, are almost never curtailed and very secure compared to water from the projects described in the previous paragraph.
- The fourth method is to lease surface water that derives from one of the previous two sources. The majority of such leases have a one-year term, meaning that the lease price is essentially the spot price of water over a yearly frequency. So, a late entrant without surface water rights or a contract with one of the major projects could, theoretically, meet their water needs through the leasing market. Practically, however, Ferguson (2025) documents that the surface water market is very illiquid, in part due to steep costs of physical settlement. Moreover, even when traded, the price of one-year surface water lease contracts is very volatile and depends on drought severity in California, as shown in Appendix Figure A.3 and mentioned in the text.

Concerned with potential groundwater depletion, the state legislature passed the Sustainable Groundwater Management Act (SGMA) in 2014. The SGMA which introduced a framework for establishing groundwater sustainability goals that does not need to be met until the early 2040s.

Tradability of Groundwater.

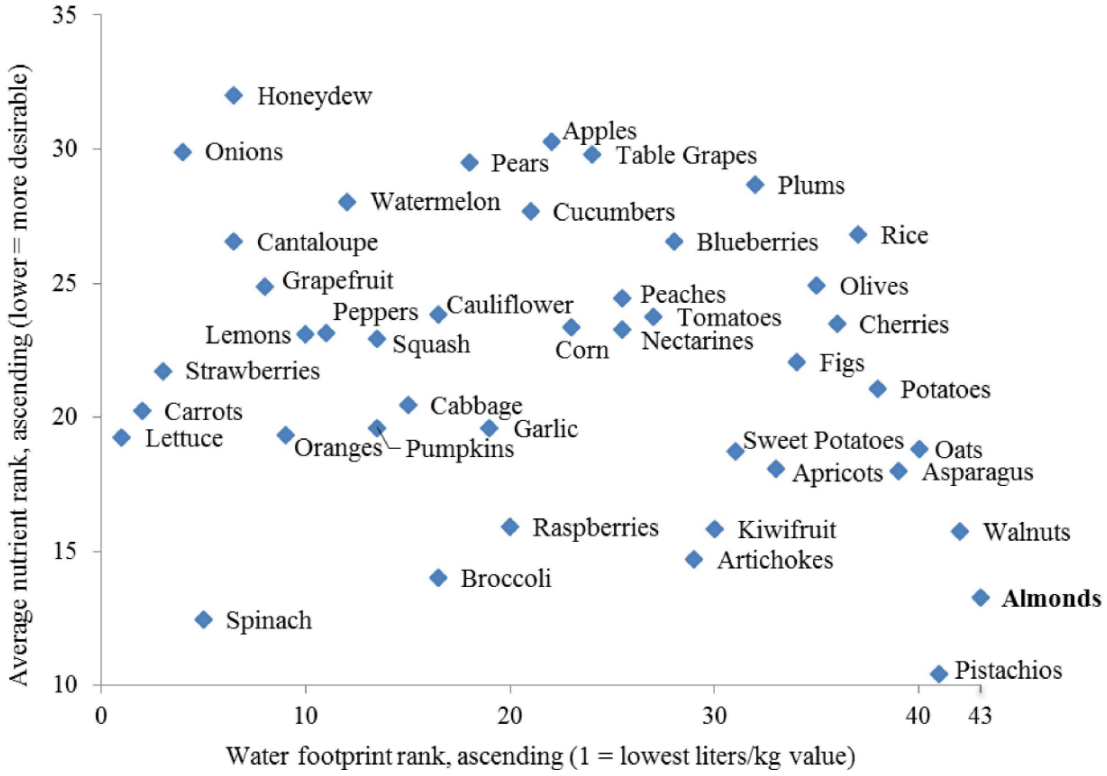
We noted in the text that groundwater is non-traded in our sample. Here, we clarify that groundwater actually can be traded in so-called adjudicated basins, which are a subset of all basins in the state and mostly located in the greater Los Angeles area. However, the statement in the text about non-tradability is practically correct in the context of tree nut cultivation, as only 0.03% of 2023 tree nut cropland is in adjudicated basins.

Note on Tree Nut Futures from Section 8.

In Section 8, we noted the absence of futures markets for almonds, pistachios, or walnuts in the text. Here, we clarify that there were two attempts to establish futures markets that, practically, would not have been available to growers in our sample. The Shenzhen Commodity Exchange attempted to introduce almond futures contracts in 2013, but, to the best of our knowledge, the Shenzen contract was never traded after 2013. The Bombay Stock Exchange attempted to introduce almond futures contracts in 2020, but the Bombay contract stopped trading as of 2025.

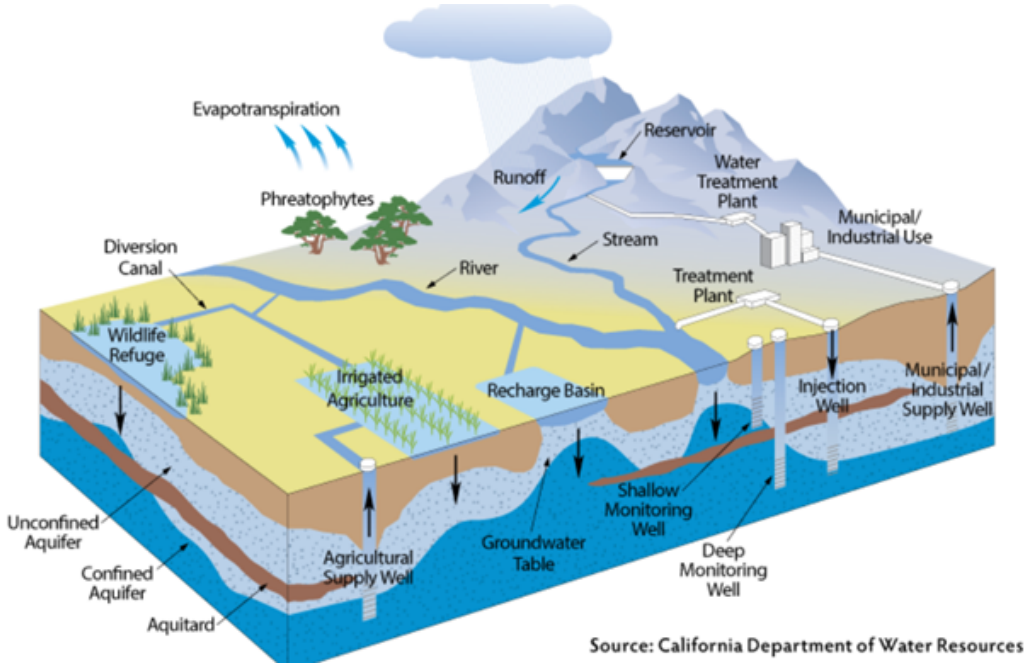
Additional Figures and Tables

Figure A.1: Water Requirements by Nutritional Value.



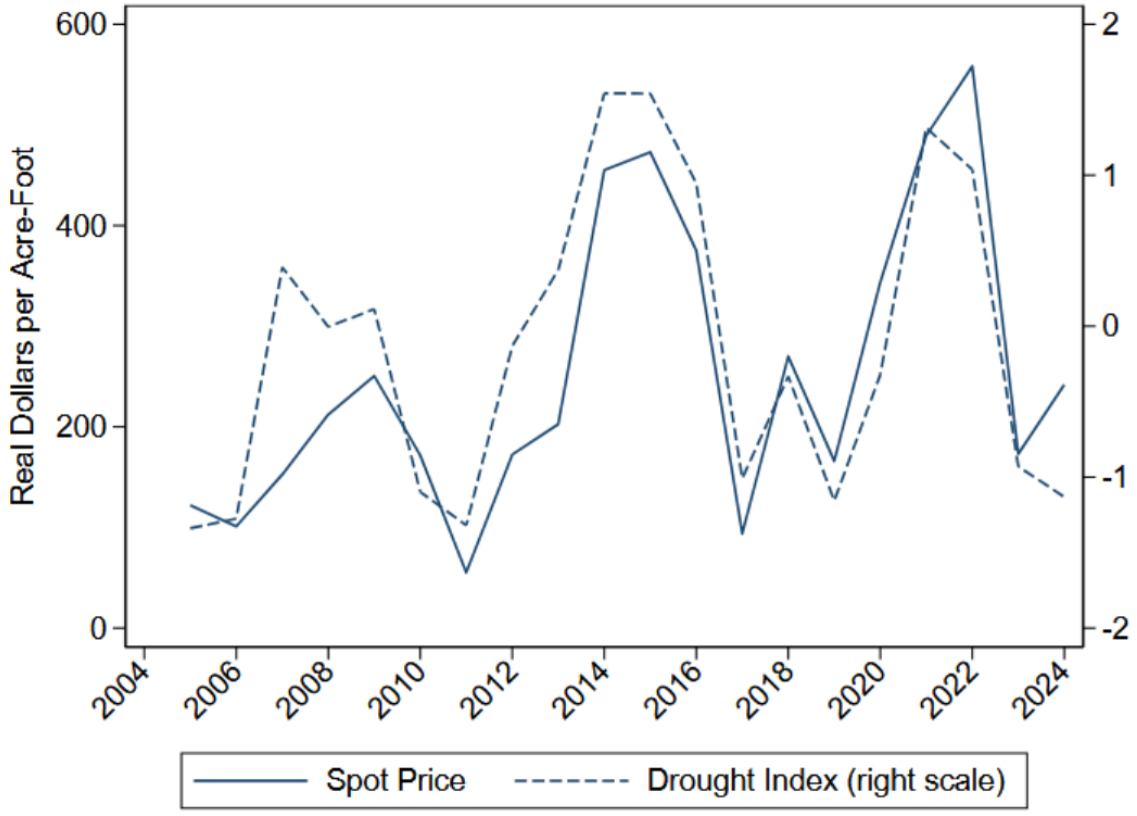
Note: This figure corresponds to Figure 3 of Fulton, Norton and Shilling (2019), and it plots the relation between water requirements and nutritional value for selected crops. The exact figure was obtained from the associated public presentation slides in Fulton (2021), not the published article.

Figure A.2: Schematic of Groundwater Basin.



Note: This figure presents a stylized schematic of a groundwater basin from the California Department of Water Resources.

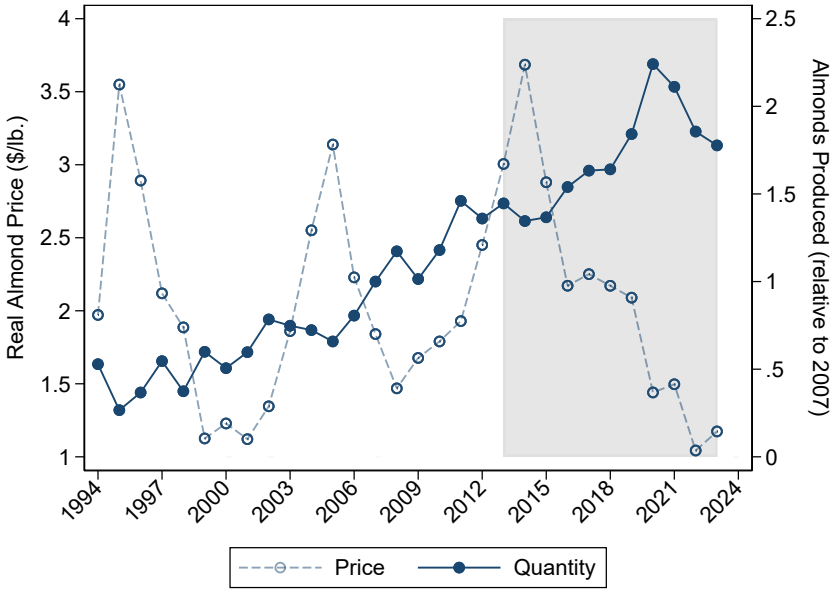
Figure A.3: Spot Price of Surface Water.



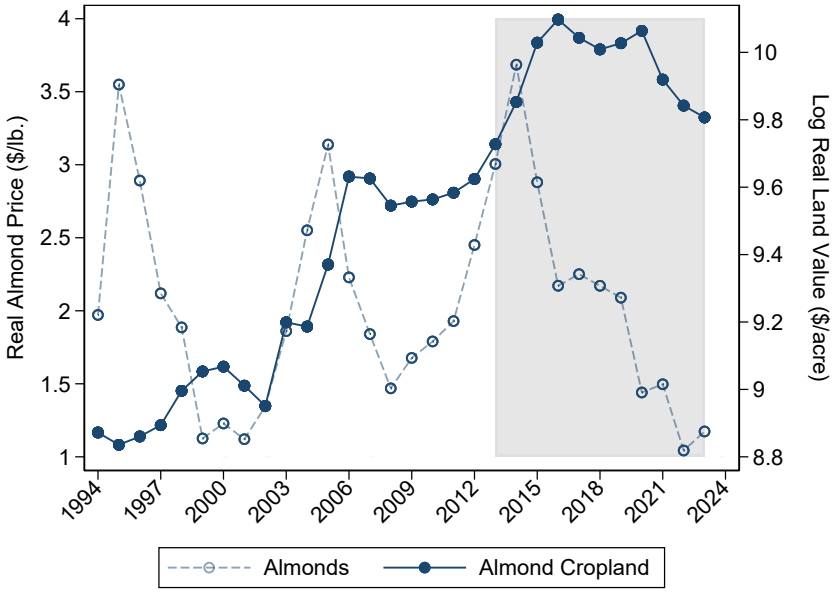
Note: This figure plots the average volume-weighted price of surface water leases of duration less than or equal to one year, based on transaction-level data from Waterlitix. Prices are deflated by CPI. The Drought Index corresponds to the California Drought Severity Index from the National Drought Mitigation Center at the University of Nebraska-Lincoln, standardized to have unit variance and mean zero.

Figure A.4: Almond Quantities and Prices and Value of Almond Cropland.

(a) Quantity of Almonds Produced



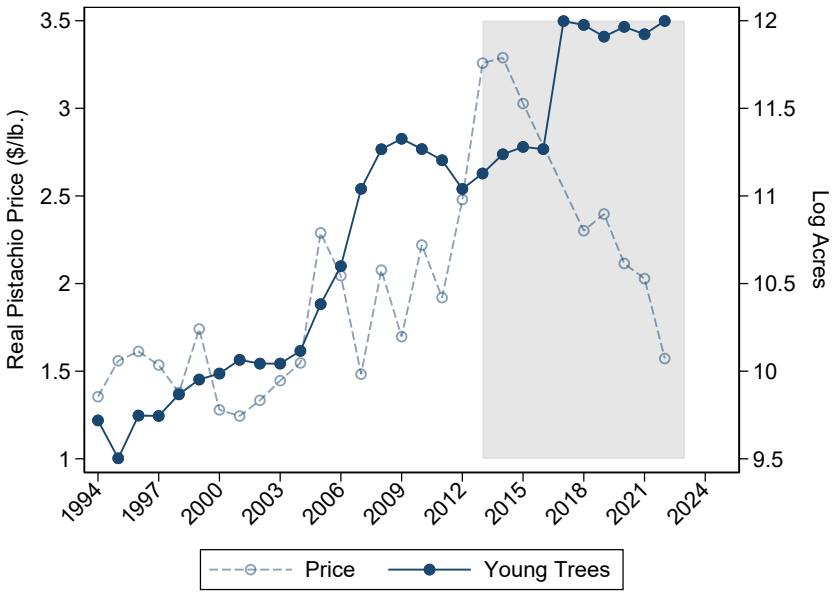
(b) Value of Almond Cropland



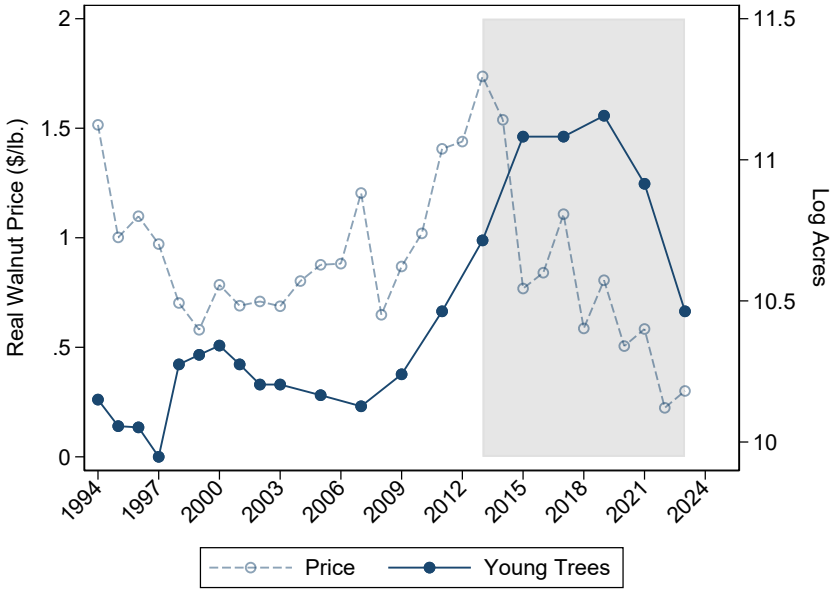
Note: This figure is analogous to Figure 2 except that it also includes quantity of almonds produced and the real appraised value of almond cropland, which is designed to approximate the market price of almond cropland. Both nut prices and land values are in real terms (2010 dollars), based on total CPI. Data on cropland appraisals are from the NCREIF data set and so represent land owned by fiduciaries that manage money for large tax-advantaged institutions. The remaining notes are the same as in Figure 2.

Figure A.5: Real Pistachio and Walnut Prices and Investment in New Trees.

(a) Pistachios

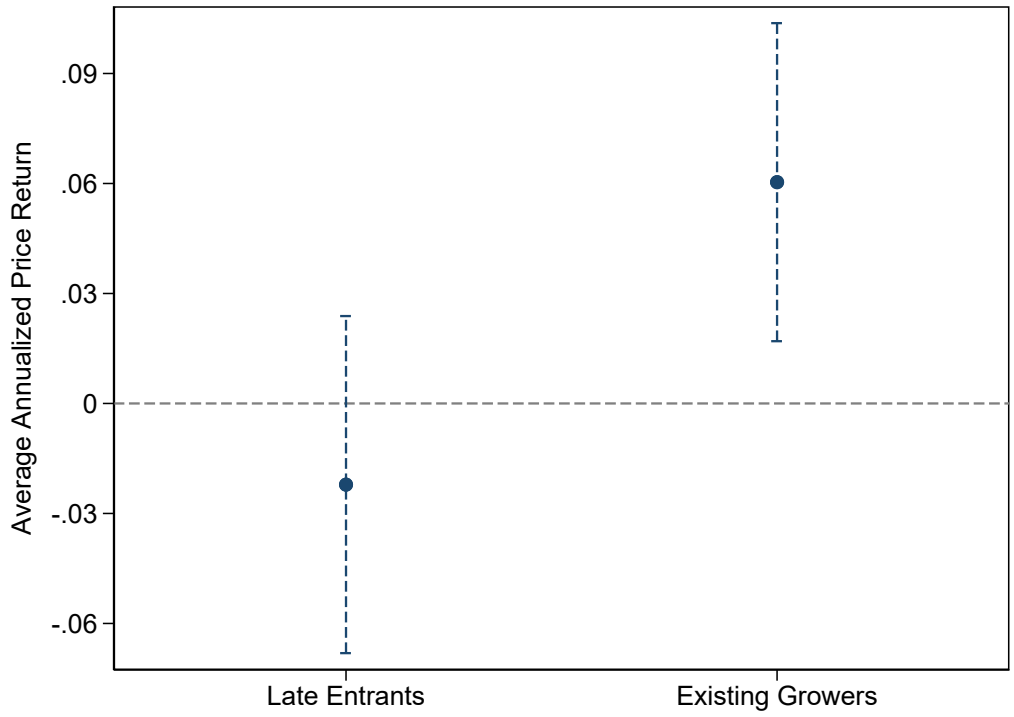


(b) Walnuts



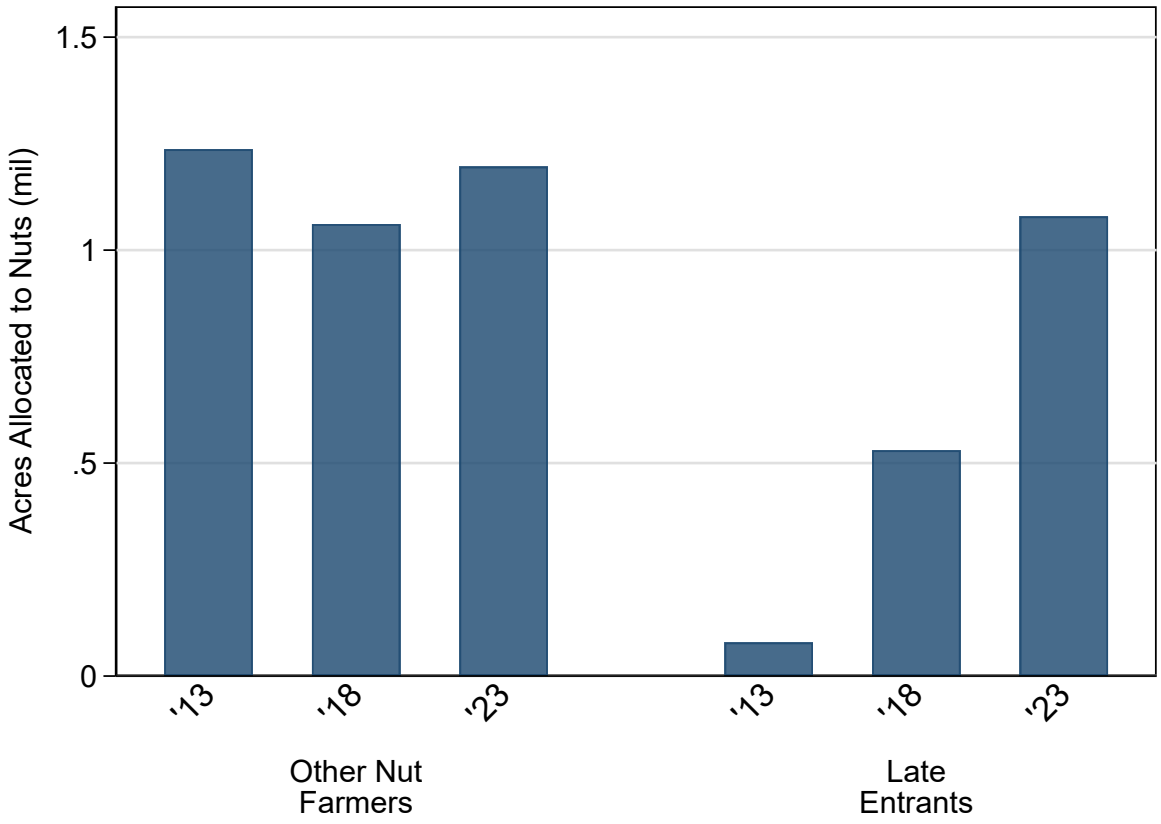
Note: This figure is the same as Figure 2 except in terms of pistachios and walnuts, as it plots real prices and acres of young trees (i.e., non-bearing acres) for these other two nuts. As in Figure 2, acres of young nut trees approximate tree investment. Data on walnuts are from the USDA California Walnut Acreage and Objective Measurement Reports. Data on pistachios are from the Administrative Committee for Pistachios Processors' Producer Delivery Reports and Acreage Surveys because the USDA does not produce current reports for pistachios. The remaining notes are the same as in Figure 2.

Figure A.6: Price Return on Land with Nut Growth.



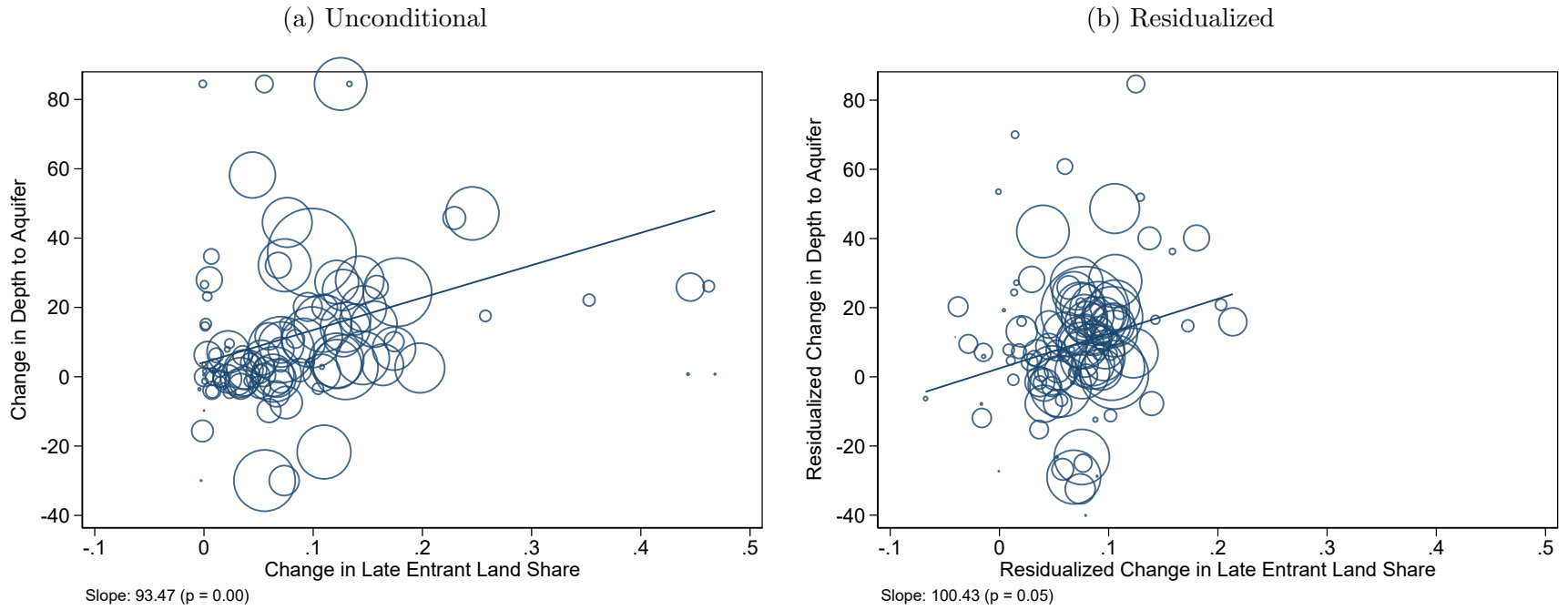
Note: The figure plots the average annualized price return on parcels sold in 2024 or 2025. The average is calculated separately for parcels owned by late entrants in the middle of the boom (2018) and for land owned by existing growers then. The set of parcels used to construct the figure are those in the raw CoreLogic-Cropscape data set after imposing the standard filters that (i.e., that the parcel have no less than half its land allocated to active or fallowed cropland in 2023 and that the parcel is not owned by the federal or state government in 2023). In addition, we restrict the sample to parcels that are sold in 2024 or 2025 and that have growth of at least 1 pp in the share of land allocated to tree nuts over 2013-2023. Annual returns are geometric and nominal.

Figure A.7: Robustness of Figure 4. Dropping Parcels Prone to Error in Name Grouping.



Note: The figure is analogous to Figure 4 except that it drops parcels smaller than 25 acres. Per the description of the name grouping algorithm in Appendix A.3, the algorithm is more likely to mis-classify individual, non-corporate growers, who plausibly own smaller plots of land, here measured as less than 25 acres. The remaining notes are the same as in Figure 4.

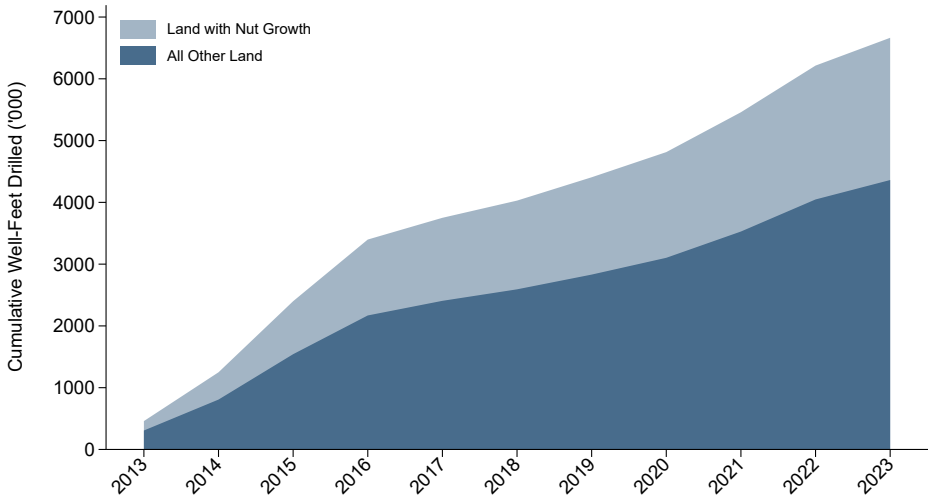
Figure A.8: DAUCO-Level Relation Between Aquifer Depth and Late Entrants.



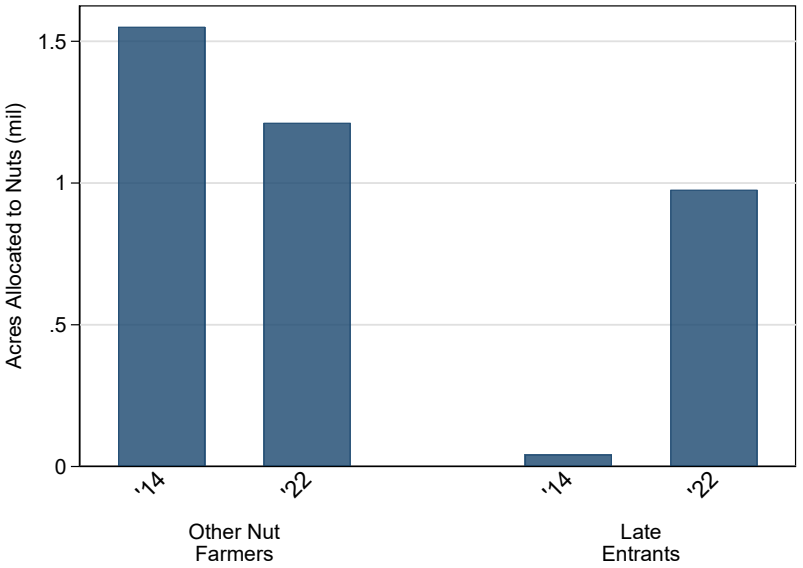
Note: This figure assesses robustness to how we construct measurement stations, which are our main unit of analysis, by plotting the relation between the 2013-2023 change in depth to the aquifer and expansion by late entrants to the nut market at the level of the detailed-analysis-unit-by-county (DAUCO). Each marker is a DAUCO, and the size of the marker corresponds to the DAUCO's area. A DAUCO is a geographic unit defined by the California DWR that intersects detailed analysis units with counties. The variable on the vertical axis is the average of $\Delta AquiferDepth_s$ across measurement wells in DAUCO d. The variable on the horizontal axis is the change in acres of nut cropland in d over 2013-2023 on land owned by late entrants in 2018, the midpoint of the boom, divided by the total acres in d, where late entrants are again defined as allocating less than 10% of their land in the CoreLogic-Cropscape data set to tree nuts in 2013. This variable is the DAUCO-level analogue of the variable $\Delta LateEntrantLandShare_s$. Panel (a) plots the unconditional relation. Panel (b) plots the relation after residualizing both the vertical and horizontal variables by DAUCO-level analogues of the following variables: $\Delta OtherNutLandShare_s$; $\Delta PerennialShare_s$, $\Delta RowShare_s$, and $\Delta FallowShare_s$ (the analogue of $\Delta NonCroplandShare_s$ is the residual); $PerennialShare_{s,13}$, $RowShare_{s,13}$, and $NutShare_{s,13}$, $FallowShare_{s,13}$ (the analogue of $NonCroplandShare_{s,13}$ is again the residual), and $AquiferDepth_{s,13}$.

Figure A.9: Aggregate Dynamics with Alternative Land Cover Data.

(a) Well Drilling by Nut Growth



(b) Nut Cropland by Entry into the Boom.



Note: This figure assesses robustness of the aggregate patterns shown in Figures 3 and 4 by replicating them using alternative land cover data from LandIQ. The LandIQ data differ from the core USDA CropScape land cover data in that we do not observe LandIQ data for 2013 and 2023, and, so, we collect data for 2014 and 2022. Panel (a) is similar to Figure 3, and it plots the cumulative number of feet of groundwater wells drilled from 2013 through 2023 on parcels in our core analysis sample with: positive growth in the share of land allocated to nuts over 2014-2022 (Land with Nut Growth); and the part from the remaining parcels in the data set (All Other Land). We do not restrict to parcels with at least 5 pp growth in this figure due to the high accuracy rates reported by LandIQ in Appendix A.5. Panel (b) is similar to Figure 4, and it plots acres of nut cropland (in millions) in a given year on parcels in our core analysis sample that were owned by existing nut growers or late entrants as of that year. Like in our main analysis, late entrants are defined as allocating less than 10% of their initial acres to nut cropland, although we now calculate this share using the LandIQ data, not the CropScape data, and, thus, must use 2014 as the initial year. We only produce panel (b) for 2014 and 2022 due to the aforementioned restriction on LandIQ sample period.

Table A.1: Largest Owners of Nut Cropland.

At the End of the Boom, 2023

A. Overall Nut Owners		B. Late Entrants	
Name	Acres	Name	Acres
Wonderful Nut Orchards	79,577	Trinitas	30,324
Farmland Reserve Inc	42,047	JG Boswell Company	24,147
Sandridge Partners	34,217	Gladstone Land Corp	12,670
Trinitas	30,324	Las Nogaleras	8,237
Assemi Farms	28,937	EW Merritt Farms	8,151
JG Boswell Company	24,147	Prudential Agricultural Investments	6,165
Manulife	22,639	AgIS Capital	5,269

In the Middle of the Boom, 2018

A. Overall Nut Owners		B. Late Entrants	
Name	Acres	Name	Acres
Wonderful Nut Orchards	64,398	JG Boswell Company	13,315
Farmland Reserve Inc	34,391	Trinitas	5,973
Sandridge Partners	25,052	EW Merritt Farms	4,805
Manulife	19,411	Orland Almonds Company	3,671
TIAA	18,386	Materra	3,175
Assemi Farms	16,432	Gladstone Land Corp	2,932
Akhavi	16,331	Rosedale Farming Group	2,681

Note: This table reports the largest owners of nut cropland as of 2023 (End of the Boom) and as of 2018 (Middle of the Boom). Panel A does so for all owners in our data. Panel B does so for the subset of owners with less than 10% of their land allocated to nuts in 2013. Owners are defined by applying the string grouping algorithm described in Appendix A to the CoreLogic Assessment Data Set, as described in Appendix A.3. For entities that own nut cropland through many subsidiaries, the Name column in the table reports an informal name of the parent company. For entities that own nut cropland through very few subsidiaries, the Name column reports an informal name of the most common subsidiary. The names shown above are not the legal names of the entities and are only meant to exemplify the types of entities that were revelant during the boom. Nut acres are calculated using the CoreLogic-Cropscape data set. The remaining notes are the same as in Table 1.

Table A.2: Parcel-Level Summary Statistics.

	Mean	Std. Deviation	Median
$WellFeetPerAcre_i$	0.827	2.966	0
$LateEntrant_{o(i)}$	0.64	0.48	1
$HasSurfaceWater_i$	0.211	0.408	0
$\log(TotalWater_{i,13\text{-}23})$	6.693	0.32	6.757
$SellInBust_i$	0.076	0.264	0
$PerennialShare_{i,13}$	0.15	0.288	0.005
$RowCropShare_{i,13}$	0.329	0.378	0.103
$FallowShare_{i,13}$	0.128	0.248	0.015
$NonCroplandShare_{i,13}$	0.205	0.261	0.079
$NutShare_{i,13}$	0.189	0.308	0.015
$\Delta PerennialShare_i$	-0.046	0.187	0
$\Delta RowCropShare_i$	-0.098	0.319	0
$\Delta FallowShare_i$	-0.051	0.226	-0.002
$\Delta NonCroplandShare_i$	-0.077	0.242	-0.013
$\Delta NutShare_i$	0.273	0.312	0.11
Number of Parcels: 51,200			

Note: Panel (a) summarizes variables at the parcel level, i. The sample consists of parcels in the CoreLogic-Cropscape data set that have growth in the share of land allocated to nuts over 2013-2023. The variable $LateEntrant_{o(i)}$ indicates if the owner of the parcel in 2018, indexed by $o(i)$, is a late entrant to the nut boom, defined as allocating less than 10% of their land to tree nuts in 2013 The variable $WellFeetPerAcre_{i,13\text{-}23}$ is the number of well-feet drilled per acre over 2013-2023 on i for non-measurement purposes. The variable $SellInBust_i$, indicates if i was sold during the bust period of 2024-2025. The variable $HasSurfaceWater_i$ is an indicator for whether i lies within the boundary of an entity with post-1914 rights to surface water (e.g., a water district with rights to a river or reservoir), based on data from Hagerty (2022) that draws on information from the Electronic Water Rights Information Management System and other sources. The variable $TotalWater_{i,13\text{-}23}$, is the average annual evapotranspiration on parcel i over 2013-2023, in liters per square meter, based on data from Open ET. Evapotranspiration equals the total amount of water lost through the combination of evaporation and absorption by vegetation (i.e., transpiration), and it is a common measure of consumptive water use as described in the text. The variables $PerennialShare_{i,13}$ through $NutShare_{i,13}$ are the 2013 shares of land on i allocated to the indicated land cover, and the variables $\Delta PerennialShare_i$ through $\Delta NutShare_i$ are the 2013-2023 changes in share of land allocated to these covers. Parcels are weighted by area. Details are in Section 2 and Appendix A.

Table A.3: Effects on Aquifer Depth by Type of Late Entrant.

	$\Delta Aquifer Depth_s$	
	(1)	(2)
$\Delta LateEntrantLandShare_s \vert Purchased$	38.298***	
	(7.043)	
$\Delta LateEntrantLandShare_s \vert NotPurchased$	49.021***	
	(6.654)	
$\Delta LateEntrantLandShare_s \vert NewToAgriculture$		43.621***
		(6.846)
$\Delta LateEntrantLandShare_s \vert NotNewToAgriculture$		46.775***
		(6.056)
$\Delta OtherNutLandShare_s$	32.131***	32.292***
	(6.377)	(6.438)
Station Controls	Y	Y
R-squared	0.242	0.240
Number of Observations	1,336	1,336

Note: This table estimates separate effects of late entrants' expansion by the type of late entrant. The specification is similar to that in column (2) of Table 3, except that $\Delta LateEntrantLandShare_s$ is decomposed into mutually exclusive components. In column (1), $\Delta LateEntrantLandShare_s$ is divided into land owned by a late entrant in 2018 but by a different owner in 2013 ($\Delta LateEntrantLandShare_s|Purchased$) and land that the late entrant owned in both 2018 and 2013 ($\Delta LateEntrantLandShare_s|NotPurchased$); In column (2), $\Delta LateEntrantLandShare_s$ is divided into land owned by late entrants in 2018 who had zero ownership of agricultural land in California in 2013 ($\Delta LateEntrantLandShare_s|NewToAgriculture$) and land owned by late entrants who had positive ownership of California agricultural land in 2013 but allocated less than 10% of it to tree nuts ($\Delta LateEntrantLandShare_s|NotNewToAgriculture$). The remaining notes are the same as in Table 3.

Table A.4: Dropping Small Parcels Prone to Error in Name Grouping.

	$WellFeetPerAcre_s$		$\Delta AquiferDepth_s$	
	(1)	(2)	(3)	(4)
$\Delta LateEntrantLandShare_s$	1.907***	1.880***	40.841***	42.985***
	(0.423)	(0.448)	(5.492)	(5.898)
$\Delta OtherNutLandShare_s$	1.139***	1.037***	28.690***	29.628***
	(0.318)	(0.301)	(6.985)	(8.360)
Size Cutoff (Acres)	25	50	25	50
Station Controls	Y	Y	Y	Y
R-squared	0.154	0.139	0.241	0.242
Number of Observations	1,283	1,152	1,283	1,152

Note: This table assesses the scope for bias from errors in the name grouping algorithm by re-estimating the main specifications in column (2) of Tables 2 and 3 after first dropping parcels smaller than 25 acres in columns (1) and (3) and smaller than 50 acres in columns (3) and (4). Per the description of the name grouping algorithm in Appendix A.3, the algorithm is more likely to mis-classify individual, non-corporate growers, who plausibly own smaller plots of land, here measured as less than 25 or 50 acres. These thresholds are chosen to match how corporate buyers account for 46% and 51% of transactions above each respective threshold over 2000-2023. The remaining notes are the same as in Tables 2 and 3.

Table A.5: Using Alternative Data on Land Cover Classifications.

	$WellFeetPerAcre_s$	$\Delta AquiferDepth_s$
	(1)	(2)
$\Delta LateEntrantLandShare_s$	1.500***	18.718***
	(0.231)	(6.013)
$\Delta OtherNutLandShare_s$	0.578*	-1.464
	(0.296)	(3.166)
Station Controls	Y	Y
R-squared	0.16	0.229
Number of Observations	1,336	1,336

Note: This table assesses the scope for bias from errors in classifying land covers in the baseline USDA CropScape data by re-estimating the main specifications in column (2) of Tables 2 and 3 using alternative land cover data from LandIQ. The independent variables $\Delta LateEntrantLandShare_s$ and $\Delta OtherNutLandShare_s$ are calculated using the 2014-2022 changes in the share of s allocated to tree nuts on land that was owned by late entrants and by other growers in 2018, respectively. We calculate the difference over 2014-2022, not 2013-2023, because we do not observe land cover in the LandIQ data for 2013 and 2023. As in the main analysis, late entrants are defined as allocating less than 10% of their initial acres to nut cropland, although we now calculate this share using the LandIQ data, not the CropScape data, and, thus, must use 2014 as the initial year. The station controls are the same as in column (2) of Tables 2 and 3, except that the initial and contemporaneous changes in land covers are calculated using the LandIQ data. We also cannot include fallow cropland as a separate category, since the LandIQ data do not distinguish between fallow cropland and other non-cropland. The remaining notes are the same as in Tables 2 and 3.

Table A.6: Allowing for Heterogeneous Treatment Effects across the Controls.

	$WellFeetPerAcre_s$	$\Delta AquiferDepth_s$
	(1)	(2)
$\Delta LateEntrantLandShare_s$	1.452**	41.525***
	(0.610)	(11.260)
$\Delta OtherNutLandShare_s$	0.905	28.244***
	(0.553)	(6.100)
Station Controls	Y	Y
R-squared	0.175	0.252
Number of Observations	1,336	1,336

Note: This table assesses whether estimating equation (1) with controls yields different estimates than the uncontrolled specification because including controls changes the weights across the heterogeneous treatment effects that comprise β. Specifically, the table re-estimates the specifications in column (2) of Tables 2 and 3 after including interactions between $\Delta LateEntrantLandShare_s$ and all the other station controls, including $\Delta OtherNutLandShare_s$. The station controls and $\Delta OtherNutLandShare_s$ are re-centered to have means of zero. The remaining notes are the same as in Tables 2 and 3.

Table A.7: Alternative Normalizations to Account for Tree Density and Age.

	$WellFeetPerNormalization_{i,\text{13-23}}$			
	(1)	(2)	(3)	(3)
$LateEntrant_{o(i)}$	3.254***	1.012**	1.822***	1.710***
	(0.411)	(0.412)	(0.627)	(0.632)
Normalization	$NutAcres_{i,23}$		$\Delta NutAcres_{i,\text{13-23}}$	
Parcel-Level Controls		Y		Y
Station FE	Y	Y	Y	Y
R-squared	0.338	0.349	0.336	0.350
Number of Observations	51,200	51,200	51,200	51,200

Note: This table assesses whether differences in tree density or age between late entrants and existing growers drive the parcel-level results from Table 6. The specification is the same, except that the number of well-feet drilled over 2013-2023 is now normalized by: acres of tree nut cropland as of 2023 in columns (1)-(2); and the change in acres allocated to tree nuts over 2013-2023 in columns (3)-(4). Normalizing by acres of nut cropland accounts for the possibility that late entrants drill more simply because they plant more densely. Normalizing by the change in tree nut acreage accounts for the fact that young trees require less water than mature trees (e.g., Yaghmour et al. 2016; Brar et al. 2015; Hasey et al. 2015). As in Table 6, the sample restricts to parcels with growth in the share of land allocated to nuts over 2013-2023. The coefficients differ between the controlled and uncontrolled specifications in part because the denominators of the dependent variables are linear combinations of the 2013 land covers and the 2013-2023 changes in these covers. The remaining notes are the same as in Table 6.

Table A.8: Robustness to Defining the Middle of the Boom as 2021.

	$WellFeetPerAcre_s$	$\Delta AquiferDepth_s$
	(1)	(2)
$\Delta LateEntrantLandShare_s$	1.685***	45.736***
	(0.530)	(6.098)
$\Delta OtherNutLandShare_s$	1.019*	30.842***
	(0.599)	(5.801)
Station Controls	Y	Y
R-squared	0.154	0.241
Number of Observations	1,336	1,336

Note: This table assesses robustness to measuring the middle of the boom as 2021, versus 2018 in the main specifications. That is, $\Delta LateEntrantLandShare_s$ is the change in acres of nut cropland over 2013-2023 on land owned by late entrants in 2021 divided by the total acres in s, where, as in the main specification, late entrants are defined as allocating less than 10% of their land in the CoreLogic-Cropscape data set to tree nuts in 2013. The analogous change in the share of land in s allocated to tree nuts on parcels not owned by late entrants in 2021 is denoted $\Delta OtherNutLandShare_s$. The remaining notes are the same as in Tables 2 and 3.

Table A.9: Robustness of Aggregated Effect.

	Assumed Replacement Share, λ	
	$\lambda = 1$ (Inelastic Demand)	$\lambda = 0$ (Elastic Demand)
Increase in Aquifer Depth, 2013-2023 (feet)		
Actual	6.11	6.11
Counterfactual without late entrants		
... based on Table A.5(2)	5.15	5.22
... based on Table A.12(1)	5.26	2.83
Share of increase due to late entrants (η)		
... based on Table A.5(2)	16%	15%
... based on Table A.12(1)	14%	54%

Note: This table is analogous to Table 8, and it reports the contribution of new nut entrants to the increase in aquifer depth (i.e., decline in groundwater levels) over 2013-2023 based on different estimates of equation (1) shown in the table. We modify the calculation shown in equation (6) when using the estimates from Table A.12(1) to account for how they obtain from a log specification. The remaining notes are the same as in Table 8.

Table A.10: Using Nut Cropland Valuations to Forecast Returns.

	$TotalReturn_{n,t\to t+3}$		$IncomeReturn_{n,t\to t+3}$	
	(1)	(2)	(3)	(4)
$\log(LandValuePerAcre_{n,t})$	-0.331*		-0.138	
	(0.185)		(0.103)	
$CapitalizationRate_{n,t}$		4.856**		1.967*
		(2.056)		(1.008)
Nut Type FE	Y	Y	Y	Y
R-squared	0.120	0.083	0.075	0.049
Number of Observations	48	48	48	48

Note: This table forecasts returns to the nut portfolio of institutional investors using the valuation of nut cropland, which assesses whether market participants overestimate the persistence of nut booms. Subscripts n and t denote nut type and year. Returns are aggregated from the quarterly to the yearly level to account fo seasonality. The data include returns for two nut types, almonds and pistachios. All columns include nut type fixed effects. The independent variables are: the log of the value of real (i.e., deflated by CPI) nut cropland per acre for nut type n in year t, denoted $\log(LandValuePerAcre_{n,t})$; and the ratio of net operating income to the value of nut cropland, denoted $CapitalizationRate_{n,t}$. Valuation of nut cropland comes from appraisals of professionals associated with the American Society of Farm Managers and Rural Appraisers and is intended to represent market value. The outcomes are the total return and the income return on the portfolio of nut type n from year t to $t+3$. Returns are geometric. The sample period is 1992 through 2023. Standard errors are Driscoll-Kraay and allow for cross-sectional correlation in a given year and auto-correlation with a lag of $T^{1/4} \approx 2$ years.

Table A.11: Assessing Externalities via Property Price Returns.

	$PriceReturn_i$
	(1)
$\Delta LateEntrantLandShare_s$	-0.140**
	(0.069)
$\Delta OtherNutLandShare_s$	0.087
	(0.158)
Hedonic Controls from Table 4	Y
R-squared	0.046
Number of Observations	810

Note: This table estimates the effect of late entrants' expansion on non-late entrants' property price returns in the same station, which assesses externalities generated by late entrants' expansion. The regression equation is of the form

$$PriceReturn_i = \beta \Delta LateEntrantLandShare_{s(i)} + \gamma Z_{s(i)} + u_i$$

where i and s index parcel and station, respectively, and $PriceReturn_i$ is the annualized nominal price return on parcel i. Specifying the outcome as a price return absorbs time-invariant property characteristics. The sample is restricted to parcels that sell over 2024-2025, excluding parcels that both were owned by a late entrant in 2018 and experienced growth in the share of land allocated to tree nuts over 2013-2023. The hedonic controls in $Z_{s(i)}$ are the same as in Table 4. Data on sales over 2024-2025 are from Regrid. Standard errors are clustered by station. The remaining notes are the same as in Tables 2 and 4.

Table A.12: Miscellaneous Robustness.

	$\Delta \log(AquiferDepth_s)$	$\Delta AquiferDepthPrePeriod_s$			$TotalWater_{i,\text{13-23}}$
	(1)	(2)	(3)	(4)	(5)
$\Delta LateEntrantLandShare_s$	0.677***	8.582	8.049	4.188	
	(0.092)	(5.819)	(8.825)	(4.828)	
$\Delta OtherNutLandShare_s$	0.518***	7.898	2.021	7.729	
	(0.081)	(5.123)	(4.761)	(6.251)	
$LateEntrant_{o(i)}$					5.217*
					(2.901)
Unit of Analysis	Station	Station	Station	Station	Parcel
Pre-Period		'02-'12	'08-'12	'02-'10	
Station Controls	Y	Y	Y	Y	
Parcel Controls					Y
Station FE					Y
R-squared	0.160	0.214	0.044	0.205	0.794
Number of Observations	1,329	777	960	773	51,200

Note: This table reports the results of various robustness tests mentioned in the paper. Column (1) is analogous to column (2) of Table 3 after replacing the outcome with the change in log depth to the aquifer over 2013-2023, and, correspondingly, replacing the initial depth to the aquifer with the initial log depth to the aquifer in the set of controls. Columns (2)-(4) are analogues of column (2) of Table 3 after replacing the outcome with the change in depth to the aquifer over various pre-periods indicated in the table, which serves as a placebo test and as an evaluation of pre-trends. Column (5) is analogous to column (3) of Table 7 after replacing the outcome with the level of average annual evapotranspiration on parcel i over 2013-2023, in liters per square meter.