Investment Booms and Natural Resources *

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

May 2026

Abstract

We document a novel inefficiency of real investment booms: over-investment can deplete scarce natural resources, like groundwater. Our empirical setting is the mid-2010s boom in California tree nuts and associated use of groundwater. The boom is driven by late entrants, who, relative to existing growers, grow on cheaper land, have shorter holding periods, are less geographically-concentrated, and are more likely to tap groundwater for irrigation. These patterns suggest that late entrants under-internalize their impact on groundwater depletion. We find that areas more exposed to late entrants experienced substantially greater declines in groundwater levels and more well drilling during the boom. We estimate that over-investment by late entrants accounts for roughly one-fifth of the total decline in groundwater levels in our sample of agricultural California over 2013-2023.

Keywords: Commodities, cropland, entry, investment booms, investors, groundwater, water, well drilling.

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

*We thank participants at the Almond Alliance's Annual Convention for practical insights. 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, IE Business School. 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

Investment often follows a boom-bust pattern (e.g., Hayek 1931; Kydland and Prescott 1982). In a typical pattern, expectations of higher future profits incentivize firms to invest in capital today, leading to an investment boom and subsequent bust once expectations are revised down again. It is well-known that private investment decisions during the boom can generate pecuniary externalities or otherwise induce spillovers to the real economy (e.g., Greenwald and Stiglitz 1986). This paper empirically shows that investment booms can also generate an environmental externality: the depletion of scarce natural resources through excessive capital accumulation.

Our idea is conceptually quite simple. Suppose that capital investment requires a rival, non-excludable resource. For example, investing in electric car batteries requires lithium, and investing in perennial cropland requires water. The key friction is that private investment decisions do not internalize how production of capital can deplete the commonly-shared resource. Consequently, real investment booms can affect the stock of this resource. Our main contribution is to empirically substantiate this idea in a particular boom-bust episode using cross-sectional variation in the degree to which firms internalize the common resource.

Our specific empirical laboratory is the mid-2010s boom-bust cycle in California tree nuts and the associated depletion of groundwater, a natural resource used to cultivate the nut trees. Section 3 describes four reasons why this laboratory is ideal for studying our motivating question of how real investment booms affect natural resources.

1.) Groundwater scarcity presents an existential threat to communities throughout the world. For example, Jasechko et al. (2024) document a global decline in groundwater levels over the past 45 years that has accelerated in the past two decades. In many cases, the decline in groundwater levels has been large enough to prompt real changes in economic activity. For example, the recent introduction of data centers in water-stressed regions of India has been associated with out-migration of local residents as scarce groundwater is allocated towards the data centers (e.g., Inamda 2025; Krishnamurthy 2025). In Iran, groundwater scarcity has been listed as an underlying driver of the 2025-2026 protests (e.g., Shokri 2026). In our setting of agricultural California, a dry and crop-intensive region, groundwater levels have steadily and significantly fallen over the past 45 years, as shown in Figure 1.

2.) Planting California tree nuts is a long-duration, water-intensive, and geographically-concentrated investment. We focus on three specific nuts – almonds, pistachios, and walnuts – of which California is a dominant global supplier due to specific advantages in climate and soil. Nut trees require three to six years from planting until they can bear fruit, thus substantiating our interpretation of them as capital. Moreover, cultivating the three particular nuts that we study requires more water per unit of nutrition than other crops, thus substantiating our assumption that investment depends on a natural resource. During droughts, like that which California experienced in the mid-2010s, growers shift from surface water to groundwater, which was then unregulated and scarce. This substantiates our assumption that the resource (i.e., groundwater) is rival and non-excludable.

3.) Tree nut investment follows recurring boom-bust cycles. In particular, the mid-2010s boom begins with a sharp increase in nut prices over 2013-2015 due to the combination of growing global nut demand and a drought in California that restricts nut supply. As with investment booms more generally, this increase in prices is shortly followed by an increase in investment (i.e., tree planting), as growers expect the elevated prices to persist. However, in the early 2020s, prices collapse amid an oversupply of nuts and trees. This same general boom-bust pattern has repeated itself roughly every decade since at least the early 1990s.

4.) Nearly all of the net expansion of tree nut cropland in the mid-2010s came from late entrants to the boom, defined as allocating very little of their pre-boom portfolio to tree nuts. These late entrants generally behave in a way that is inconsistent with them planning to stay in this market for the long-term.[1] For example, the average late entrant in our data plants nut orchards on land that was formerly open space, has erodible soil, has a large pre-existing depth to the aquifer, and is otherwise of low-quality, based on a hedonic valuation. Moreover, relative to existing nut growers, late entrants are more likely to drill unregulated groundwater wells and less likely to acquire long-term surface water rights.

[1] As an illustrative example, the largest late entrant in our sample, a private equity fund named Trinitas Partners, sold off its portfolio and entered bankruptcy 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).

Using this setting as a laboratory, we pursue our more-general question of how investment booms affect natural resources by specifically estimating the effect of late entrants to the nut boom on the change in groundwater levels over 2013-2023. We do so using a rich data set, described in Section 4, that draws on multiple sources and tracks ownership, transactions, land cover, soil characteristics, water consumption, and well drilling at the parcel level in agricultural California over 2013-2023. We also observe groundwater levels, measured in depth to the aquifer, at various measurement wells managed by the California Department of Water Resources. We conduct most of our analysis at the level of stations, which are small geographies that collect parcels with the same measurement well, and, thus, approximately the same depth to the underlying aquifer (Bierkens and Wada 2019).

Our research design, described in Section 5, compares changes in groundwater levels in stations with more versus less expansion by late entrants to the nut boom over 2013-2023. We can interpret this difference as the causal effect of late entrants absent two potential forms of bias. The first potential form of bias is reverse causality. Pure reverse causality would bias our estimates towards zero, since, if anything, late entrants would have an incentive to expand where groundwater levels are expected to decline less. The second form of bias concerns omitted variables that jointly predict changes in groundwater levels and where late entrants expand. We address this concern by controlling for a host of variables such as: the initial distribution of land cover and changes in it; soil characteristics; exposure to drought and regulation; pre-trends in groundwater levels; and the change in existing growers' nut cropland, which aggregates unobserved variables relevant for tree nut planting.

In Section 6, we estimate that greater exposure to late entrants' nut expansion corresponds to a significantly greater increase in depth to the aquifer over 2013-2023. This effect is weaker, though still quite significant, where late entrants expand by endogenously purchasing new land as opposed to developing land they already own, consistent with selection biasing the estimate towards zero. In addition, stations where late entrants expand more experience significantly greater well drilling over our 2013-2023, which suggests that the relative decline in groundwater levels reflects more-intensive reliance on wells by late entrants.

To better understand the scope for omitted variables bias, we examine patterns in where late entrants expand. As referenced earlier, late entrants expand more in stations with non-cropland, erodible soil, and a lower initial groundwater level. Informally, these patterns suggest that late entrants expand on cheap and low-quality land, perhaps because they do not plan to stay in the market for the long-term. Indeed, a hedonic regression based on market valuation supports this interpretation that late entrants expand on cheap land.

Interestingly, late entrants are also significantly more likely to expand on land close to tree nut processors, and, yet, proximity to such processors does not command a significant market price in the hedonic valuation.[2] These two observations, respectively, suggest that instrumenting for late entrants' expansion using proximity to a nut processor would satisfy the relevance condition and the exclusion restriction of a 2SLS estimator. Accordingly, our 2SLS estimates imply that exposure to late entrants corresponds to twice the decline in groundwater levels as implied by the baseline OLS estimate. Given the strong first stage, this larger effect may, again, reflect conservative bias in our baseline estimates.

An exhaustive set of robustness tests described in Section 7 further supports this conservative interpretation. These tests include: replicating our results using alternative satellite data on land cover (LandIQ) that draws on different sources than our baseline data from the USDA; verifying the absence of significant pre-trends in groundwater levels; replicating our station-level results at a more-aggregated geographic unit (DAUCO); dropping parcels more subject to measurement error in land ownership; and correcting for how adding controls changes treatment weights (e.g., Angrist 1998).

Next, we verify that late entrants drill more intensely at the parcel-level, which provides strong support for the validity of the main results as it holds fixed all unobserved station-level characteristics through a station fixed effect. Our parcel-level analysis also sheds light on the margins through which expansion by late entrants lowers groundwater levels by more relative to existing nut growers. In particular, our evidence suggests that late entrants are less likely to expand on land with surface water rights, a valuable yet costly-to-acquire alternative to groundwater. Moreover, they consume more water overall than existing nut growers. This may reflect how existing growers choose to irrigate less than the static optimum because

[2]Practically, being close to a processor reduces the risk that nuts develop mold or alfatoxins before they can be suitably dried-out at a processor (e.g., U.S. Environmental Protection Agency 2025; Chen and Pan 2022). It also facilitates access to nut-specific labor and harvesting technologies, such as tree shakers, that can congregate in the vicinity of a processor.

they wish to preserve their water sources for future use, whereas late entrants apply as much water as needed because they do not expect to stay in the market for the long-term.

We have performed our analysis under the tacit interpretation that late entrants internalize groundwater depletion less than existing nut growers. Several pieces of evidence in Section 8 substantiate this interpretation, beyond the fact that, by definition, late entrants did not have a stake before the boom. First, late entrants are more likely than existing growers to sell their land during the bust of 2024-2025, which implies that they have a shorter holding period. Intuitively, a grower with a shorter holding period has less incentive to preserve groundwater for future use as they will not be in the market then, which aligns with the logic of Hotelling (1931). Second, the average late entrant owns a smaller share of both the station and the groundwater sub-basin in which its land is located, relative to other owners. This result suggests that late entrants may perceive their own groundwater use to have small effect on the overall aquifer, as in the original Tragedy of the Commons.

Section 9 concludes by using use our estimates to approximate the aggregate, in-sample effect of late entrants' nut investment on groundwater levels. Specifically, we compare the actual decline in groundwater levels to the decline that would have occurred in a counterfactual without late entrants. This aggregation exercise requires an assumption about the geometry of the aquifer and an assumption about the elasticity of demand for tree nuts, which affects how existing growers would behave under the counterfactual. Since this elasticity must be bounded between zero and infinity, we can also obtain bounds on the effect of late entrants' investment on aggregate groundwater levels. Our results imply that investment by late entrants to the nut boom accounts for around one-fifth of the overall decline in groundwater levels in our sample of agricultural California over 2013-2023.

Overall, our results point to a newly-documented externality through which investment booms can affect the real economy. Unlike pecuniary or aggregate demand externalities, this natural resource externality arises because of a constraint on supply imposed by nature. So, the externality could potentially become stronger over time as natural resources become more scarce and nature's constraint begins to bind more tightly. In particular, natural resources like groundwater that depend on the weather for natural recharge may become increasingly scarce in the presence of climate change.

Related Literature

Our main contribution is to document what is, to the best of our knowledge, a novel externality of real investment booms. A large macro-finance literature has theoretically characterized how private real investment can generate pecuniary or aggregate demand externalities or otherwise induce capital misallocation.[3] By contrast, the externality that we document − 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. Empirically, a recent set of papers has documented recurring boom-bust 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 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.[4]

Second, we contribute to a broad literature on free entry by providing new empirical evidence that unrestricted entry can generate inefficiencies. In the context of real estate investment, our focus on the economic effects of new entrants in tree nut cropland parallels a quickly-growing literature on the economic effects of entrants in out-of-town residential real estate markets (e.g., Chinco and Mayer 2016; Badarinza and Ramadorai 2018; Gao, Sockin and Xiong 2020; Cvijanović and Spaenjers 2021; Favilukis and Van Nieuwerburgh 2021; Gorback and Keys 2025). Unlike these other real estate papers, we study a different economic effect of free entry, namely, that it exacerbates a common pool problem. At a conceptual level, this idea is essentially a variant of Levhari and Mirman (1980), who show

[3] A very partial set of 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).

[4] 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.

theoretically how reducing competition can incentivize firms that rely on a common pool to extract less over time.[5]

Third, we contribute to a recent empirical literature on the economics of groundwater. We particularly complement Hadachek et al. (2024), who also study how shocks to growers in agricultural California affect groundwater levels over recent decades. While Hadachek et al. (2024) 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.

These findings naturally raise the question of whether and how policymakers should regulate groundwater. Early theoretical work by Gisser and Sanchez (1980) shows how, if there is enough storage capacity in the aquifer, then the private and social planner extraction levels approximately align, which would imply a modest role for policy. Practically, California policymakers have indeed begun to regulate groundwater through the Sustainable Groundwater Management Act (SGMA), which requires local jurisdictions to construct and implement plans for groundwater management by the early 2040s, thus incentivizing anticipatory extraction beforehand (Bruno and Hagerty 2025). Our results suggest that the composition and, in particular, the holding period of groundwater users would be a relevant input for policy discussion. Moving beyond our particular setting, the natural resource costs of real investment booms may provide a novel rationale for macroprudential policies.

2 Theory

The theory we have in mind begins with the idea that capital investment requires a rival, non-excludable resource. We specifically have in mind that firms use the resource to build capital, as in the case of, say, lithium used to build electric cars or water used to cultivate

[5] 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 radio broadcasters (e.g., Berry and Waldfogel 1999) or real estate agents (e.g., Hsieh and Moretti 2003; Barwick and Pathak 2015). The opposite extreme of a public monopoly is also inefficient, as shown empirically by, for example, Seim and Waldfogel (2013) and Verboven and Yontcheva (2024).

perennial cropland. The logic would be similar if, instead, the natural resource is used as a complement to capital in production, as in the case of water to cool data centers, although the timing of resource depletion would then be shifted into the future.

Since capital takes time to build, forecasts of higher future profits incentivize firms to invest more in capital today. We call this pattern of expectations-induced investment "a boom". Our theory does not depend on whether firms make forecasts using rational expectations, although systematic over-reaction to recent profit would amplify the main effect.

The key friction in our theory is that private investment decisions do not internalize how production of capital can deplete the commonly-shared resource. Such non-internalization can arise if firms have a high effective discount rate, and, thus, have little incentive to preserve the resource for their own future production. Or, it can arise in markets with small, competitive firms, who, by virtue of their small size, perceive little impact on the resource. When empirically evaluating the theory, we utilize cross-sectional heterogeneity in the degree of non-internalization. We particularly focus on late entrants to the boom, who, due to their short holding periods and limited exposure to any single common pool, may internalize the social costs of resource extraction less than other firms.

The depletion of the natural resource during the boom could potentially exacerbate other well-documented inefficiencies of investment booms referenced in the introduction. For example, resource scarcity could potentially induce capital misallocation, as firms in all industries that rely on the resource scale back. We do not seek to quantify such spillovers in our empirical analysis, and, instead, pursue the more specific effects of over-investment in the boom due to non-internalization of the common resource.

A small model in Appendix B.1 formalizes the points made in this section.

3 Setting

Our empirical setting is the mid-2010s boom in almonds, pistachios, and walnuts, which we simply refer to as tree nuts. We describe why this setting is ideal for testing the theory above. Some of this discussion relies on data sets described soon in Section 4. We aim to be succinct and defer additional details to Appendix B.3.

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., Yaghmour et al. 2016; Brar et al. 2015; Hasey et al. 2015). Because of this time-to-build feature, planting nut trees is essentially a capital investment.

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.

Growers Expect Elevated Nut Prices to Persist. A necessary condition for our theory is that firms expect the current-period price of output to persist into future periods.[6] Otherwise, they would have no incentive to make capital investments that do not pay off for at least three years. So, although spikes in nut prices do not actually persist, as just documented above, all that is required is that growers forecast these elevated prices to persist. Appendix B.2 presents evidence consistent with such a forecast.

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,

[6] Growers cannot lock-in future nut prices because there are no futures markets for almonds, pistachios, or walnuts over our sample period. Neither does the USDA's Risk Management Agency offer price insurance for these nuts, although it does offer highly standardized yield (i.e., quantity) insurance.

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 such periods of drought, leasing surface water can become prohibitively expensive (Appendix Figure A.3). Groundwater, by contrast, is non-traded in our sample. 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.

Role of Late Entrants to the Boom. Late entrants to the nut boom account for the majority of aggregate expansion of nut cropland over 2013-2023, which is a novel fact that we document in Figure 4. We define late entrants as growers that allocated less than 10% of their agricultural land to tree nuts in 2013 or did not own agricultural 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.

4 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.

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 (Section 7.4).

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 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 CDL 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 CDL 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 sub-basins 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, in units of distance, 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 wells and 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 for some of our analysis.

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 are: 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.

4.1 Summary Statistics

Table 1 provides summary statistics. Our main analysis sample consists of 1,336 groundwater measurement stations. 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 the each stage of the data merging and cleaning procedure.

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. As briefly mentioned in Section 3, late entrants are defined as land owners that allocated less than 10% of their land to tree nuts in 2013 or did not own cropland in 2013, based on the CoreLogic-Cropscape data set. 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.

5 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. Economically, we interpret the estimate as the effect of over-investment in the boom due to non-internalization of the common resource used to build capital. This interpretation seems reasonable insofar as late entrants have a shorter horizon than existing nut growers or, more generally, if they have weaker incentives to preserve a common resource that helps sustain the nut market. Multiple pieces of evidence in sections 7.1 and 8 support this view.

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: $LateEntrantLandShare_s$ 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 $OtherNutLandShare_s$ 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 measures of extraction, which lower them. We conduct inference by clustering u_s at the level of the groundwater sub-basin associated with a station, since water from recharge shocks induces within-sub-basin correlation in aquifer depth.

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

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

[7] Appendix Table A.6 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 7.

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 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, which, as we shall see shortly, explain a substantial share of variation in changes in groundwater levels; and measures of drought severity and exposure to future restrictions in pumping associated with the Sustainable Groundwater Management Act (SGMA). The note to Table 1 has precise variable definitions.

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.

6 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 7.3. 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.9 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 Section 7.7. This increase in the coefficient is also consistent with the remark from Section 5 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.9 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]

[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.

Lastly, column (6) of Table 3 separately estimates the effect of $\Delta LateEntrantLandShare_s$ from land that late entrants already owned in 2013 versus 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.

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.

7 Internal Validity

7.1 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.

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.

7.2 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,

$$\begin{aligned} \Delta LateEntrantLandShare_s = \zeta DistanceToProcessor_s + \\ \tilde{\delta}\Delta OtherNutLandShare_s + \tilde{\gamma} X_s + u_s. \end{aligned} \tag{4}$$

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 5. On the other hand, the larger estimates could simply reflect weak instruments bias. However, the first stage F-statistic of 32.61 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.

7.3 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.

We estimate the following regression equation across parcels i,

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

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, and so the comparison is between late entrants and existing nut growers within the same station.[9]

The results in Table 6 imply that late entrants drill more well-feet-per-acre than existing growers over 2013-2023, making use of the parcel-level summary statistics in Appendix Table A.2. This estimate is comparable to the station-level analogue in Table 2. 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 natural question of how existing nut growers irrigate their land if they do not rely as

[9] 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.

much on well drilling. 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 have rights to surface water. Unlike groundwater, surface water rights are regulated by the California Department of Water Resources, although, according to Hagerty (2022), they are almost never curtailed. Moreover, surface water can be leased and traded in a market, commanding a high price during times of drought (Appendix Figure A.3). Together, these two observations imply that surface water rights are a valuable asset, and the costly process of acquiring them may be interpreted as 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 indeed less likely to hold surface water rights.[10] So, part of the reason why, in Table 3, the depth to the aquifer increases by less where existing growers expand may be because these growers are more likely to own rights to surface water, and, thus do not need to rely on the common pool of groundwater.

Water Consumption

A second margin is the total quantity of water applied. Existing growers may apply less water than the static optimum because they internalize that applying more water would deplete the common pool for future growing cycles. Late entrants, by contrast, may apply as much water as needed because they do not plan to remain in the market for another cycle. 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

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

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.9). 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 heightened reliance on groundwater and overall water consumption best explains why stations with more expansion by late entrants experience more of an increase in depth to the aquifer than stations with expansion by existing growers.

7.4 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.4 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 shortly in Section 9, we obtain similar implications for aggregate groundwater levels based on the LandIQ data as with our baseline CropScape data set.[11]

This similarly strongly supports the validity of the main results and provides evidence

[11] Appendix Figure A.8 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%).

that they are not driven by measurement error in the USDA's crop classifications.

7.5 Errors in Name Grouping Algorithm

The name grouping algorithm in Appendix A.3 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.5, such under-grouping would only explain our results if owners with names more prone to under-grouping systematically own land in stations with declining aquifer levels 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.

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.[12]

The resulting estimates in Appendix Table A.3 are very similar to the baseline estimates in Tables 2 and 3. Appendix Figure A.6 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.

7.6 Robustness 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 4. To address concerns about this geographic definition, we replicate a version of Table 3 at the level of detailed-analysis-unit-by-county (DAUCO). A DAUCO is a geography defined by pairs of

[12]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.

"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 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.7 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.

7.7 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.5 are quite similar to their counterparts in Tables 2 and Table 3, reinforcing our earlier interpretation that including controls reduces omitted variables bias.

8 Internalization of the Aquifer

Our analysis to this point focuses on what late entrants do and the effects of their expansion on groundwater levels. Now we turn to the economic mechanism that explains this behavior, and, in particular, our interpretation that late entrants internalize groundwater depletion less than existing growers. The small model in Appendix B.1 helps clarify how this non-internalization mechanism relates to other candidate mechanisms, such as differences in technology or expectations, and it shows how non-internalization can operate along a spatial and an intertemporal margin. The remainder of this section tests for differences in internalization between late entrants and existing growers along these two margins.

8.1 Propensity to Sell

One notion of non-internalization is intertemporal: Growers with a short holding period have less incentive to preserve the aquifer because they may not need it in future growing cycles. Some growers may have a short holding period and so be less concerned with depleting the aquifer. We test this by predicting whether parcels owned by late entrants in 2018 are more likely to be sold during the bust period of 2024-2025. Since our CoreLogic data end in 2023, we obtain data on these more recent sales from a similar data provider, Regrid, as described in Appendix A. We estimate

$$SellInBust_i = \beta LateEntrant_{o(i)} + \gamma X_i + \alpha_{s(i)} + u_i. \tag{6}$$

Note that the unit of analysis is the land parcel, i, and so we can include station fixed effects $\alpha_{s(i)}$. The sample is again restricted to parcels with growth in nut land share over 2013-2023.

The results in Table 8 show that parcels owned by late entrants are significantly more likely to be sold during the bust than other parcels with nut growth over 2013-2023. This suggests that late entrants have a shorter holding period than do existing growers since, by definition, late entrants did not own nut cropland or owned relatively little of it before the boom. That late entrants have a shorter holding period supports our interpretation that they internalize the aquifer less than do existing growers.

8.2 Share of the Aquifer Owned

Another notion of non-internalization is spatial. Some growers may own a small share of land overlying the aquifer, and so they only perceive that their drilling has a small effect on groundwater levels. We measure spatial non-internalization in two ways. The first is grower o's share of land overlying the groundwater measurement station s. The second measure is o's share of land overlying the sub-basin $b(s)$, which allows for the fact that groundwater can flow between stations within a sub-basin over the course of time. We estimate

$$\Delta ShareOfGeography_{o,g} = \beta LateEntrant_o + \alpha_g + u_{o,g}, \tag{7}$$

where the geography g is either the station or the sub-basin.

The results in Table 9 show that late entrants own smaller shares of the stations or sub-basins in which they operate. This suggests that they have less incentive to restrict groundwater extraction to preserve overall groundwater levels, since they control a smaller share of the common pool.

9 Aggregation

We attempt to quantify the natural resource costs of investment booms. To do so, 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.

We require two additional assumptions, in addition to the identification assumption (2).

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. **Geometry of the Aquifer.** The distance from the earth's surface to the aquifer is uniform within each measurement station.

Assumption 1 recognizes that, if consumers' nut demand is partially inelastic and, importantly, if existing owners forecast that reduced future nut supply will increase future nut prices through this demand inelasticity, then existing owners would expand by more in a counterfactual without new nut entrants than they do in the data. 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 more accurate for stations that are contained within a single groundwater sub-basin, defined as a geographic unit that shares the same aquifer. Reassuringly, this condition applies to 95% of stations in our estimation sample.

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

$$\begin{aligned} CounterfactualDepth_{s,23} = -\beta_\lambda \Delta LateEntrantLandShare_s + \\ \Delta AquiferDepth_s + AquiferDepth_{s,13} \end{aligned} \tag{8}$$

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 Assumption 2, we can write the aggregate change in the volume of groundwater under the counterfactual as

$$CounterfactualGroundwaterLoss_{13\text{-}23} = \sum_s \left[CounterfactualDepth_{s,23} - AquiferDepth_{s,13}\right] \times Area_s, \tag{9}$$

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

$$ActualGroundwaterLoss_{13\text{-}23} = \sum_s \Delta AquiferDepth_s \times Area_s, \tag{10}$$

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

$$\eta = \frac{ActualGroundwaterLoss_{13\text{-}23} - CounterfactualGroundwaterLoss_{13\text{-}23}}{ActualGroundwaterLoss_{13\text{-}23}} \tag{11}$$

Table 10 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 (10), after normalizing by the number of acres in our sample to maintain the same units as the first row. 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 over-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.7). Overall, this finding provides evidence that investment booms can significantly reduce the supply of scarce natural resources like groundwater.

10 Conclusion

We document what we believe to be a novel externality of real investment booms: excess depletion of natural resources. 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. These patterns suggest that late entrants internalize groundwater depletion less than existing growers. 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 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 generate 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.

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 Com-*

puting 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.

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.

Berry, S. T. and Waldfogel, J.: 1999, Free entry and social inefficiency in radio broadcasting, *RAND Journal of Economics* **30**(3), 397–420.

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, âweâll always have parisâ: 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.

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.

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.

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.: 2024, External costs of climate adaptation: Groundwater depletion and drinking water, *Technical report*, Working paper.

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.: 2025, Transaction costs and the gains from trade in water markets, *Unpublished, Working Paper* .

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.

Hotelling, H.: 1931, The economics of exhaustible resources, *Journal of political Economy* **39**(2), 137–175.

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.

Krishnamurthy, R.: 2025, Indiaâs tech boom collides with deepening water crisis in bengaluru, *Earth Journalism Network* . Accessed: 2026-02-07.
URL: *https://earthjournalism.net/stories/indias-tech-boom-collides-with-deepening-water-crisis-in-bengaluru*

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.

Levhari, D. and Mirman, L. J.: 1980, The great fish war: an example using a dynamic cournot-nash solution, *The Bell Journal of Economics* pp. 322–334.

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.

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.

Seim, K. and Waldfogel, J.: 2013, Public monopoly and economic efficiency: evidence from the pennsylvania liquor control board's entry decisions, *American Economic Review* **103**(2), 831–862.

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.

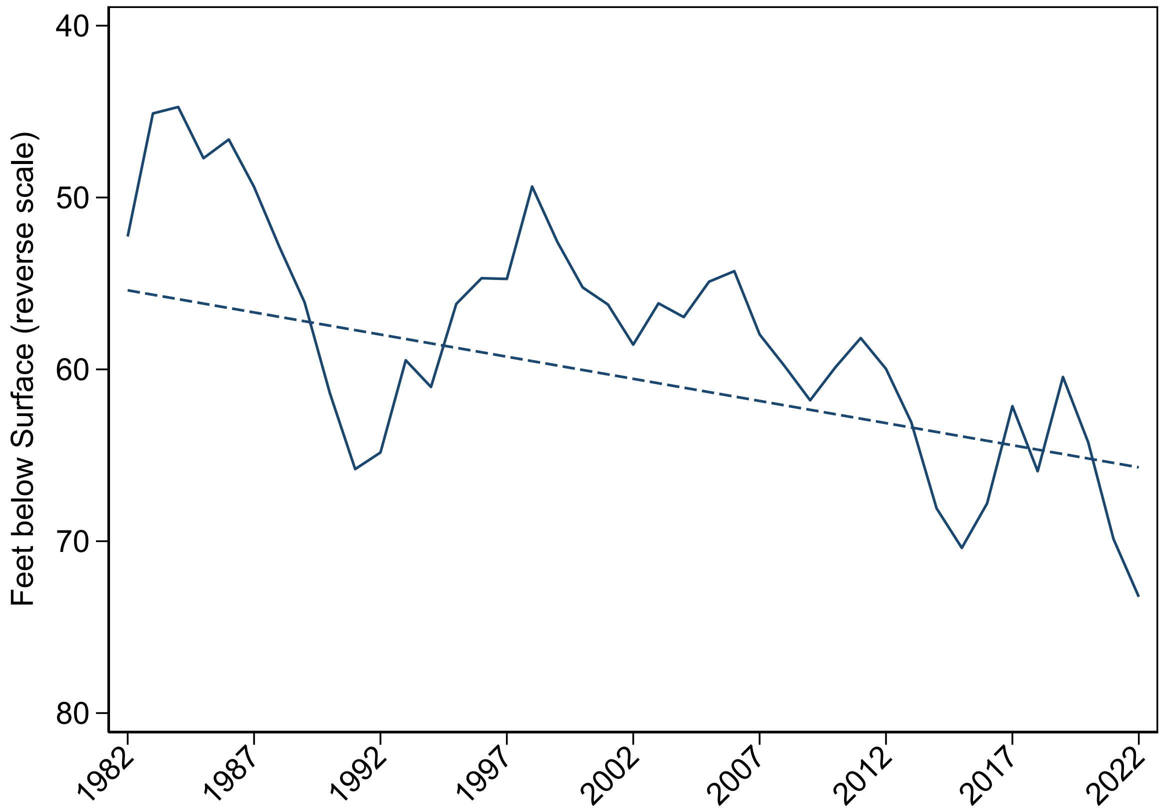
U.S. Department of Agriculture, Foreign Agricultural Service: 2025, Production, supply and distribution online (PS&D).

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.

Verboven, F. and Yontcheva, B.: 2024, Private monopoly and restricted entryâevidence from the notary profession, *Journal of Political Economy* **132**(11), 3658–3707.

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.

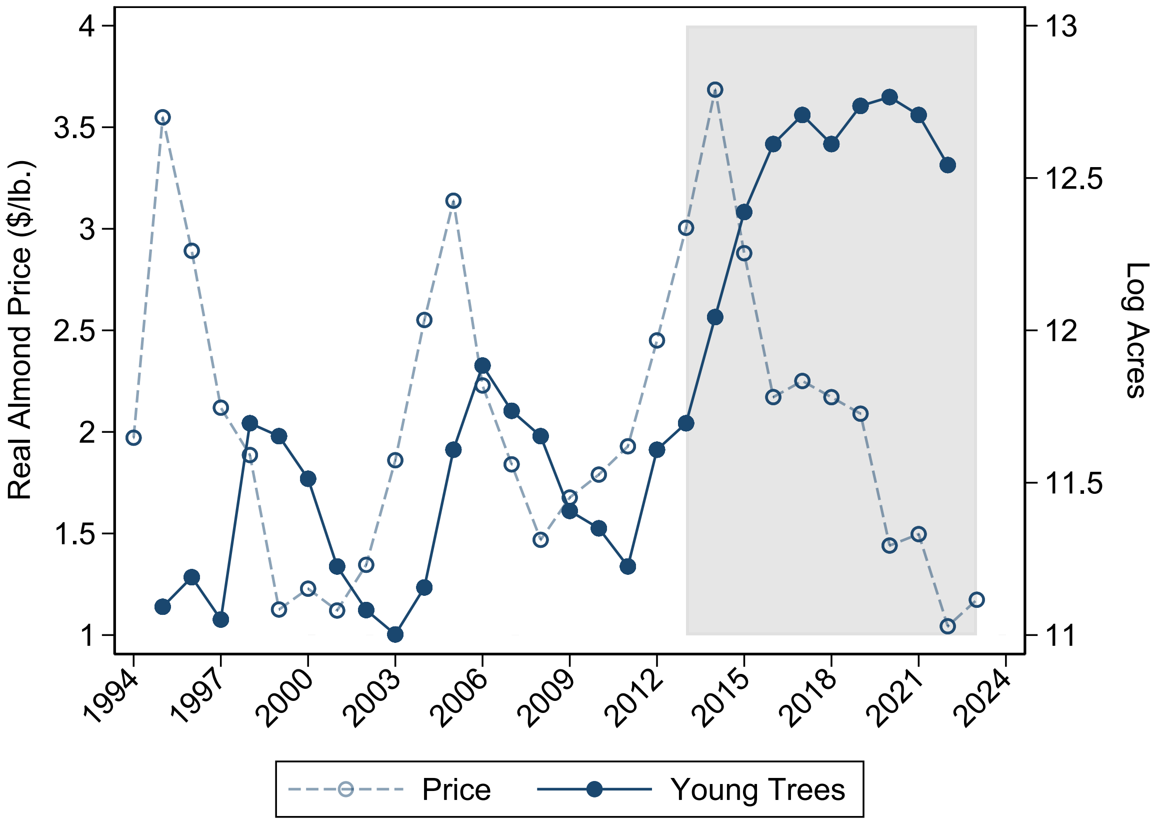


Note: The vertical axis has a reverse scale. The figure presents a time series of the depth to the aquifer across groundwater measurement stations in California. Data are from the California Department of Water Resources (DWR) Periodic Groundwater Level Measurements. To address changes in the composition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

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

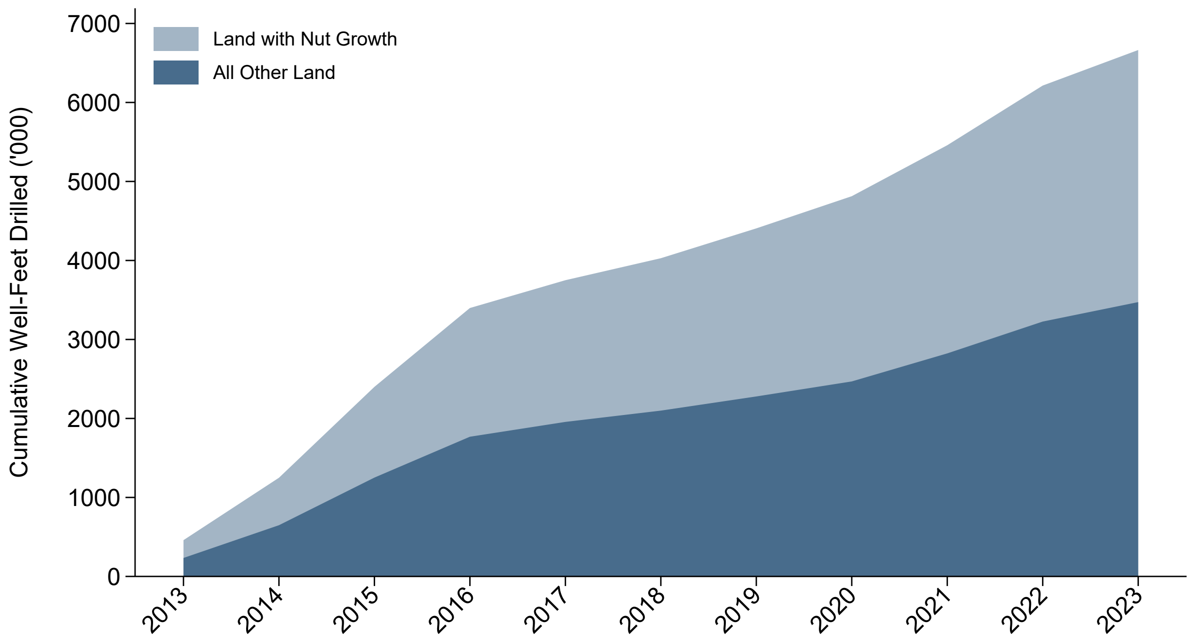
where α_w and τ_t are well and year fixed effects, respectively, and $AquiferDepth_{w,t}$ is the depth to the aquifer, in feet, as measured at w in t. The figure specifically plots the coefficients τ_t, which have the interpretation of a constant-well index. Aquifer depth is expressed in feet below land 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 4 describes our data sets.

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



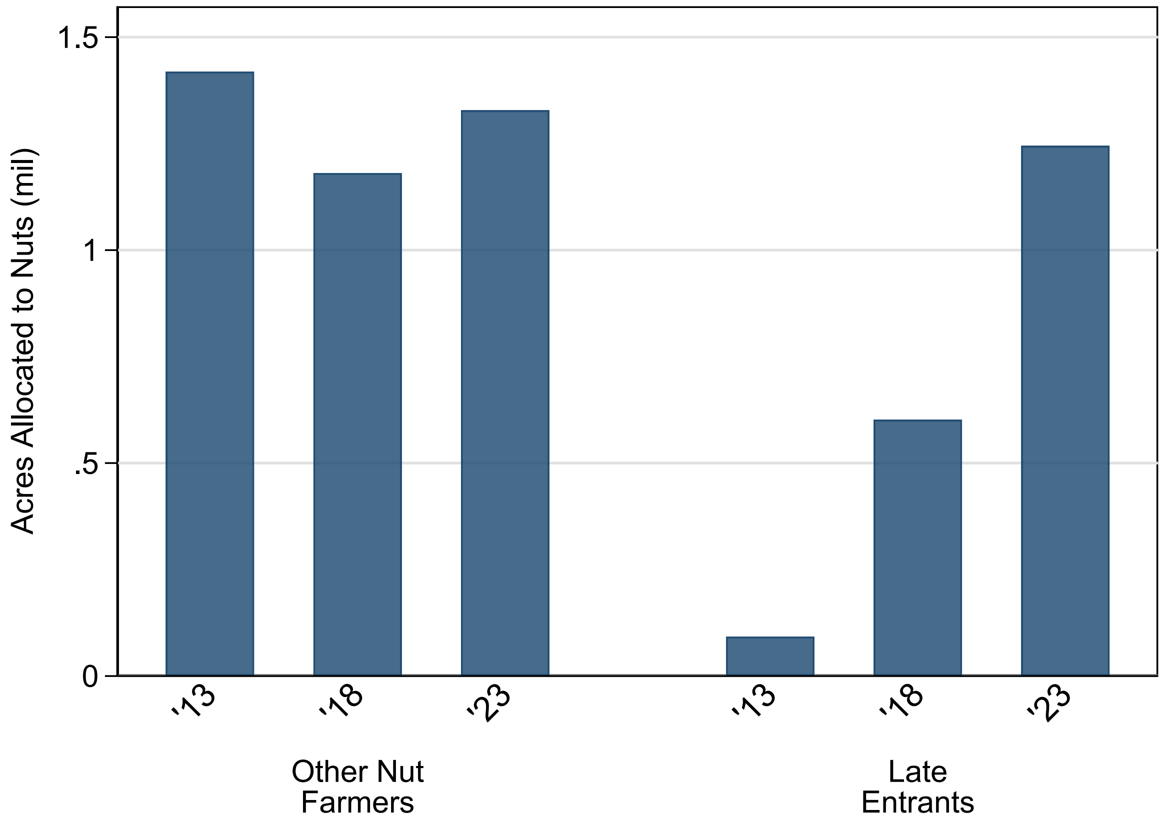
Note: This figure plots real almond prices and acres of young almond trees (i.e., non-bearing acres) in California. Since almond trees take around four years until they can begin to bear fruit, changes in the acres of young almond trees is analogous to investment in almond trees. Gray bars mark our 2013-2023 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 4 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 4 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 4 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 sub-basin, $b(s)$, 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 4 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)
<u>Controls</u>			
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: $AquiferDepth_{s,13}$ and the area-weighted station-level averages of $RowCropShare_{i,13}$, $PerennialShare_{is,13}$, $NutShare_{i,13}$, and $FallowShare_{i,13}$ from Table 1, such that $NonCroplandShare_{i,13}$ 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 sub-basin are in parentheses.

Table 3: Effect of Nut Investment on Aquifer Depth.

	$\Delta AquiferDepth_s$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta LateEntrantLandShare_s$	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***	32.131***
	(5.088)	(6.471)	(6.864)	(8.490)	(6.505)	(6.377)
$\Delta LateEntrantLandShare_s \vert Purchased$						38.298***
						(7.043)
$\Delta LateEntrantLandShare_s \vert NotPurchased$						49.021***
						(6.654)
Controls						
Crop Shares & Aquifer Depth, 2013		Y	Y	Y	Y	Y
Changes in Crop Shares, 2013-2023		Y	Y	Y	Y	Y
Soil Characteristics		Y	Y	Y	Y	Y
Drought and Regulation		Y	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	0.242
Number of Observations	1,336	1,336	777	960	773	1,336

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. Column (6) separates $\Delta LateEntrantLandShare_s$ into the component consisting of land owned by a late entrant in 2018 but by a different owner in 2013 ($\Delta LateEntrantLandShare_s|Purchased$) and that that the late entrant owned in both 2018 and 2013 ($\Delta LateEntrantLandShare_s|NotPurchased$). 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 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 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,

$$Y_s = \beta \Delta Late\widehat{EntrantL}andShare_s + \delta \Delta OtherNutLandShare_s + \gamma X_s + u_s.$$

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 Full 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 new nut 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 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: Propensity to Sell in Bust. Intertemporal Internalization.

	$SellInBust_i$	
	(1)	(2)
$LateEntrant_{o(i)}$	0.010**	0.011**
	(0.005)	(0.005)
Parcel Controls		Y
Station FE	Y	Y
R-squared	0.411	0.411
Number of Observations	51,200	51,200

Note: This table assesses whether late entrants internalize the aquifer less than existing growers because they have a shorter holding period. The unit of analysis is the land parcel, i. The regression equation has the same form as equation (2) studied in Table 6, except that the outcome, $SellInBust_i$, indicates if i was sold during the bust period of 2024-2025. Data on sales over 2024-2025 are from Regrid. The remaining notes are the same as in Table 6.

Table 9: Share of Aquifer Owned. Spatial Internalization.

	$ShareOfStation_{s,o}$	$ShareOfBasin_{b,o}$
	(1)	(2)
$LateEntrant_o$	-0.131***	-0.011***
	(0.012)	(0.002)
Standardized Outcome	Y	Y
Station FE	Y	
Sub-basin FE		Y
R-squared	0.449	0.286
Number of Observations	36,131	26,660

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 sub-basin. Each observation is an owner-by-station or owner-by-(sub)basin pair. The regression equation is of the form

$$\Delta ShareOfGeography_{o,g} = \beta LateEntrant_o + \alpha_g + u_{o,g}. \tag{12}$$

The outcomes are the share of the geography g (s for station and b for sub-basin) owned by o in 2018, the midpoint of the boom. These outcomes are then standardized to have a mean of zero and variance of one. The independent variable, $LateEntrant_o$, indicates if o is a late entrant. Standard errors clustered by owner are in parentheses.

Table 10: 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 drop in aquifer levels across groundwater measurement stations in the analysis sample, weighting by the area of each station; and the average counterfactual drop without late entrants, again weighting by area. The counterfactual in station s is computed as in equation (8) 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

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

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 9.

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 California 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 2018 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 call the resulting data set the CoreLogic-Cropscape data set.

We also make use of the CoreLogic Transaction 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 21 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.[13] 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.[14] We use this ownership information to consolidate subsidiaries of the same entity into a single owner identifier.

[13]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.

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

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 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 2013, 2014, 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 phases 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 4.

- **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 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 from 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, we drop all processors of almonds or walnuts established after 2014.

- **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, 8, and A.9). 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 Small Model

A small model helps clarify why late entrants may behave differently than existing growers. There is a unit mass of firms i. Each symmetric firm uses w_t of natural resources at time t to produce $\pi_{t+1}(w_t)$ of profit at time $t+1$. Extracting w_t units of natural resources comes at a per-unit cost $c(W_t)$, where W_t is the aggregate stock of the resource and $c(\cdot)$ is a decreasing function. The aggregate stock evolves according to

$$W_t = W_{t-1} + \omega_t - \int_i w_t(i)di, \tag{A1}$$

where ω_t is a recharge shock.

Firms optimize according to

$$V(W_t, w_{t-1}) = \max_{w_t} \left\{ \pi_t(w_{t-1}) - c(W_t)w_t + e^{-r}\mathbb{E}_t \left[V(W_{t+1}, w_t) \right] \right\}, \tag{A2}$$

where r is the discount rate. Taking the first order condition of the maximand in the Bellman equation (A2) provides intuition about the economic forces at play,

$$\underbrace{c(W_t)}_{\substack{\text{Cost of}\\ \text{Extraction}}} + \underbrace{|c'(W_t)|w_t di}_{\substack{\text{Spatial}\\ \text{Internalization}}} + \underbrace{e^{-r}\mathbb{E}_t \left[\frac{\partial V}{\partial W_{t+1}} di \right]}_{\substack{\text{Temporal}\\ \text{Internalization}}} = \underbrace{e^{-r}\mathbb{E}_t \left[\pi'_{t+1}(w_t) \right]}_{\substack{\text{Expected}\\ \text{Return}}}. \tag{A3}$$

The three terms on the left side of equation (A3) reflect private and social costs of extraction, while the term on the right side reflects the private benefit of extraction.

In more detail:

- **Private Cost.** The term $c(W_t)$ is the privately-incurred cost of extraction. Late entrants could potentially have a different extraction cost. However, such a difference is unlikely to explain our main findings given the large set of control variables, including initial depth to the aquifer, which is the empirical analogue of W_t.

- **Spatial Non-Internalization.** The term $|c'(W_t)|w_t di$ is a social cost due to the fact that firm i's private extraction raises the private cost of extraction for all firms. If each individual firm is small, such that $di \to 0$, then this social cost has little effect on an individual firm's decision. However, if all firms behave this way, then their collective decisions can non-trivially affect the stock of the resource and cost of extraction. We interpret this term as spatial internalization, since, in our empirical setting, di maps to the share of the common pool (i.e., land above the aquifer) controlled by each individual firm. The evidence in Table 9 suggests that late entrants own a smaller share of the surrounding station or sub-basin than existing growers, which would imply that they have a low value of this spatial internalization term.

- **Intertemporal Non-Internalization.** The term $e^{-r}\mathbb{E}_t\left[\frac{\partial V}{\partial W_{t+1}}di\right]$ is a dynamic social cost of extraction. Even if a firm recognizes that its private extraction affects the common pool at t, which could be modeled by inserting a mass point into the distribution of i, it may still internalize only a portion of this social cost if it plans to exit the market at $t+1$, which could be modeled through a low value of $\frac{\partial V}{\partial W_{t+1}}$. The evidence in Table 8 suggests that late entrants indeed are more likely to exit the market at the end of the bust than existing growers, which would imply a small value of this dynamic internalization term.

- **Private Benefit.** The term $e^{-r}\mathbb{E}_t\left[\pi'_{t+1}(w_t)\right]$ captures the expected $t+1$ payoff from extracting the resource at t. This expected payoff could potentially vary across growers along two margins: technology or expectations. We now relate these two margins to our empirical analysis.

  First, it is possible that late entrants have more water-intensive technologies or growing conditions than existing growers, as captured by the term $\pi'_{t+1}(w_t)$. We view parcel-level technological differences such as the use of flood versus drip irrigation as consistent with our interpretation that late entrants internalize the aquifer less than existing growers. This interpretation is reasonable insofar as such technological adoption is a long-term investment that late entrants are unwilling to make given their shorter holding period. This interpretation would not be reasonable if technological differences between late entrants and existing growers instead reflects differences in growing conditions. However, the large set of control variables used in our empirical analysis along with the robustness of the results to a parcel-level, within-station analysis as in Table 6 makes it unlikely that differences in growing conditions explain any unobserved technological differences between the two groups.

  Second, equation (A3) raises the possibility that late entrants drill more because they hold different expectations about future cash flows. However, the analysis in Appendix B.2 suggests that systematic forecasting errors can affect both existing growers and late entrants, which makes us cautious about interpreting our main results as reflecting overly-optimistic forecasts of nut prices on the part of late entrants.

B.2 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.[15]

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 4, 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.

[15] 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.8 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). 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.

B.3 Additional Institutional Details

We elaborate on the institutional details described in Section 3. We organize the additional information under the same headings used in the text.

Groundwater is a Common Pool.

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 (2025), 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.

Drought Incentivizes Use 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.

Growers Expect Elevated Nut Prices to Persist.

We noted the absence of tree nut futures markets 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.

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 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 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 Measurement Error in Late Entrant Share

We substantiate the remarks from Section 7.5 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

$$Y_s \equiv \Delta AquiferDepth_s, \tag{A4}$$

$$M_s \equiv [\Delta LateEntrantLandShare_s, \Delta OtherNutLandShare_s]. \tag{A5}$$

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

$$\widehat{M_s} = M_s + \underbrace{[\xi_s, -\xi_s]}_{\equiv \Xi_s}. \tag{A6}$$

Note the symmetry of ξ around zero in equation (A6). 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 ξs, then the same error leads us to under-estimate existing growers' expansion by ξs. The sign is reversed in the case of over-grouping.

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

$$\begin{aligned}
[\widehat{\beta}, \widehat{\delta}]' &= [\widehat{M_s}'\widehat{M_s}]^{-1}\widehat{M_s}'Y_s \\
&= [M_s'M_s + M_s'\Xi_s + \Xi_s'M_s + \Xi_s'\Xi_s]^{-1}[M_s + \Xi_s]'Y_s \\
&= \Lambda[\beta, \delta]' + \Lambda[M_s'M_s]^{-1}\Xi_s'Y_s.
\end{aligned} \tag{A7}$$

where

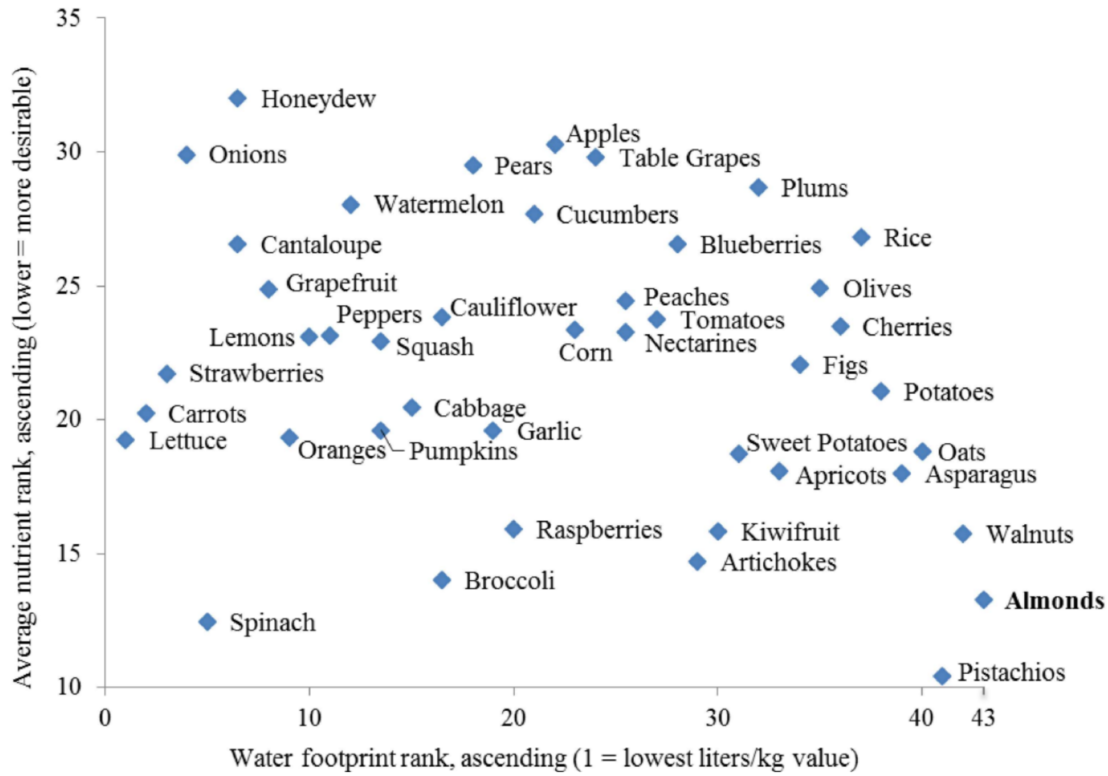
$$\Lambda = [M_s'M_s + M_s'\Xi_s + \Xi_s'M_s + \Xi_s'\Xi_s]^{-1}[M_s'M_s] \tag{A8}$$

measures the degree of attenuation bias. The second term in equation (A7) 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.

Recapping, 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.

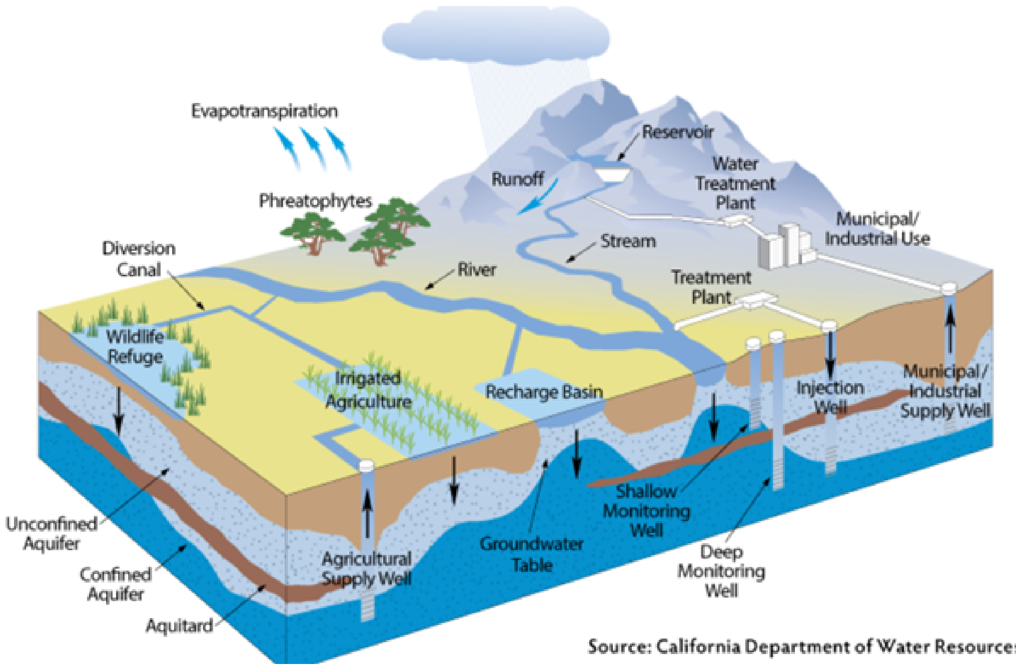
Additional Figures and Tables

Figure A.1: Water Requirements by Nutritional Value.



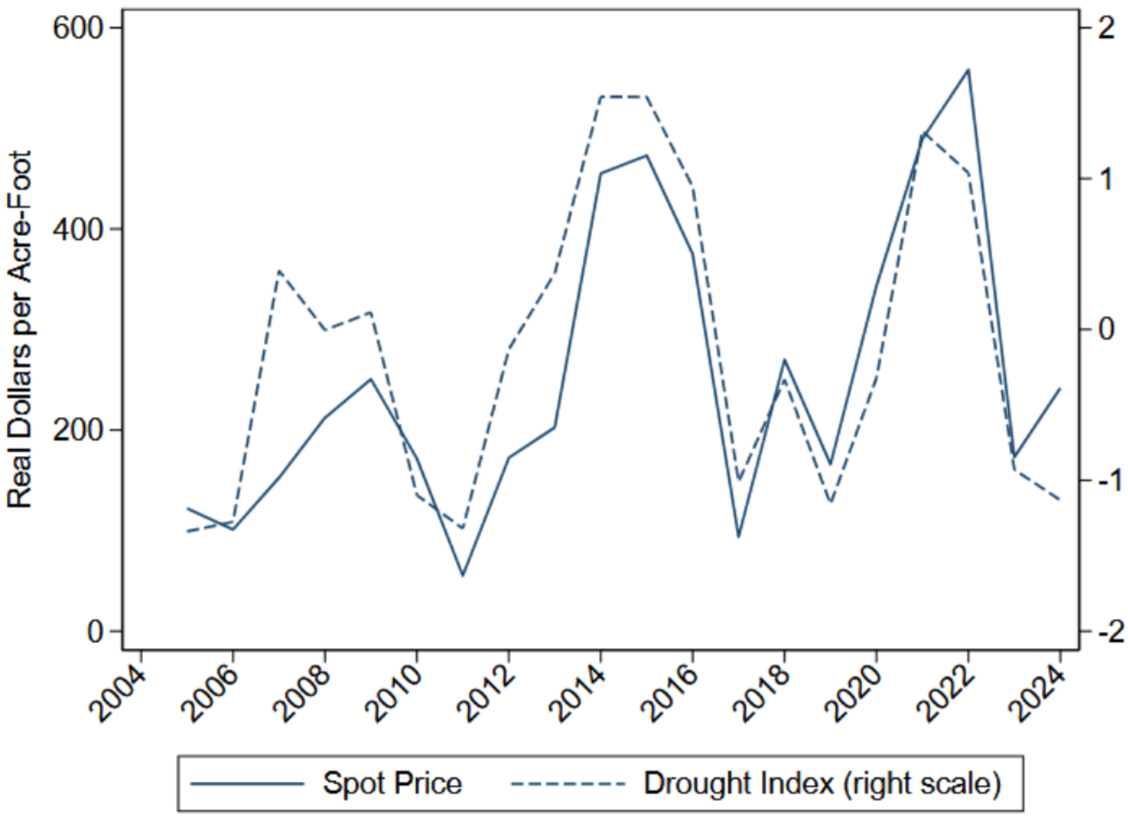
Note: This figure corresponds to Figure 3 of Fulton, Norton and Shilling (2019), and it plots the relation between water requirements and nutritional value for selected crops. The exact figure was obtained from the associated public presentation slides in Fulton (2021), not the published article.

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.

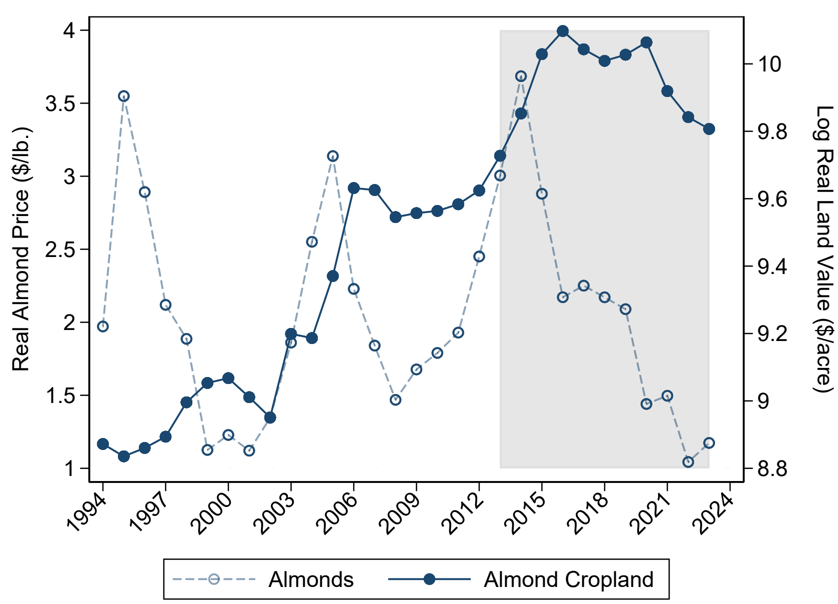


Note: This figure plots the average volume-weighted price of surface water leases of duration less than or equal to one year, based on transaction-level data from Waterlitix. Prices are deflated by CPI. The Drought Index corresponds to the California Drought Severity Index from the National Drought Mitigation 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

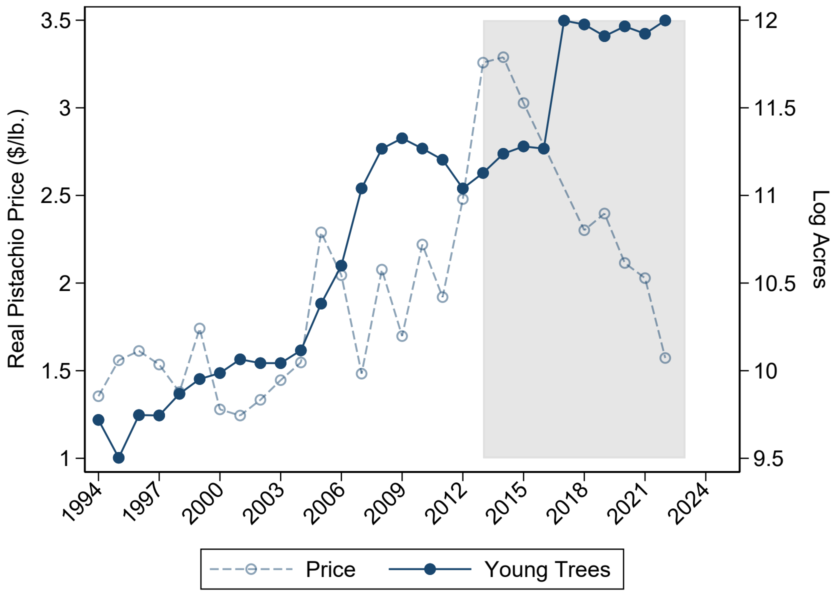
(b) Value of Almond Cropland



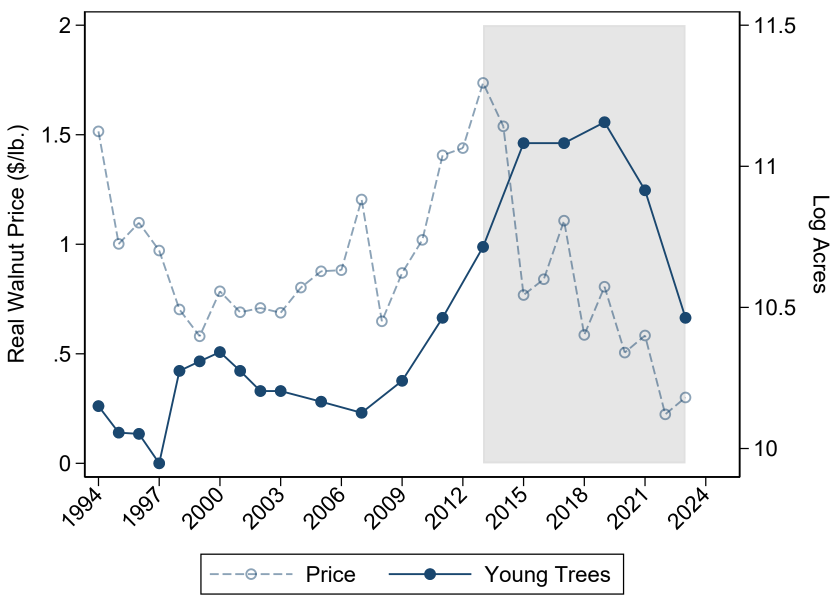
Note: This figure is analogous to Figure 2 except that it also includes quantity of almonds produced and the real appraised value of almond cropland, which is designed to approximate the market price of almond cropland. Both nut prices and land values are in real terms (2010 dollars), based on total 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

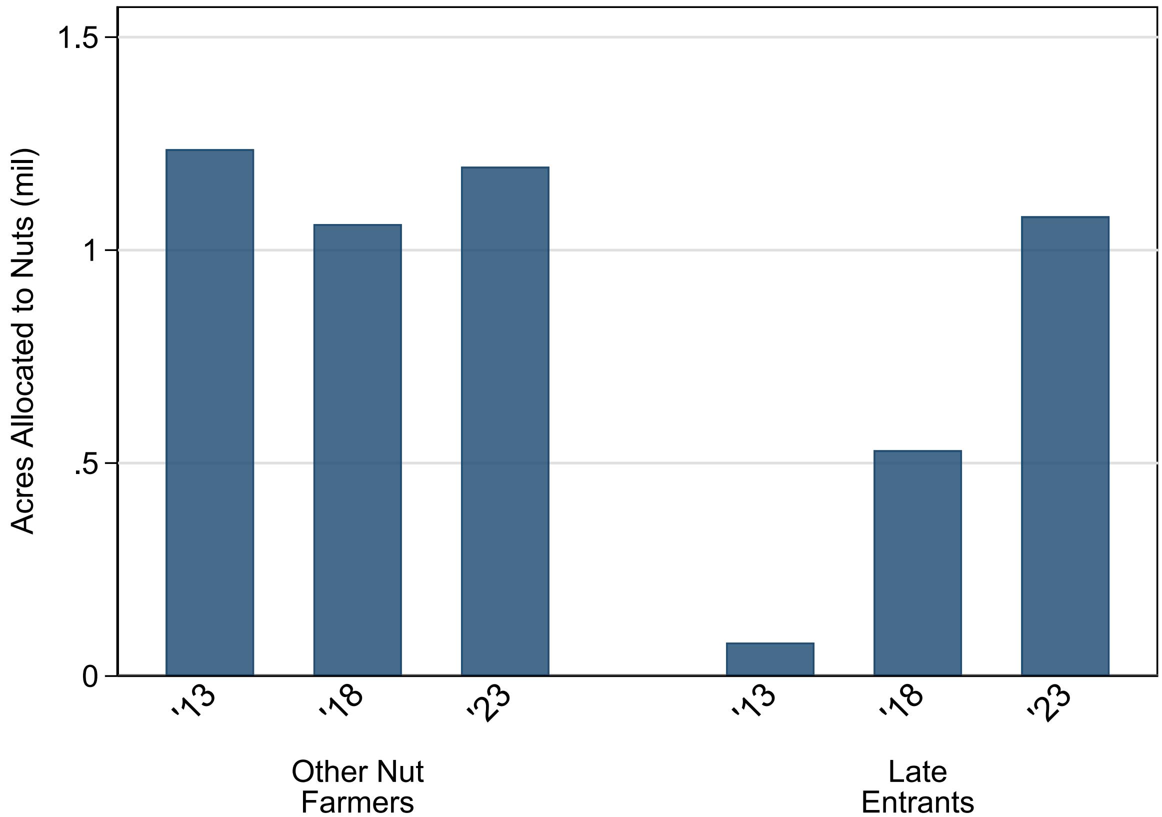


(b) Walnuts



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

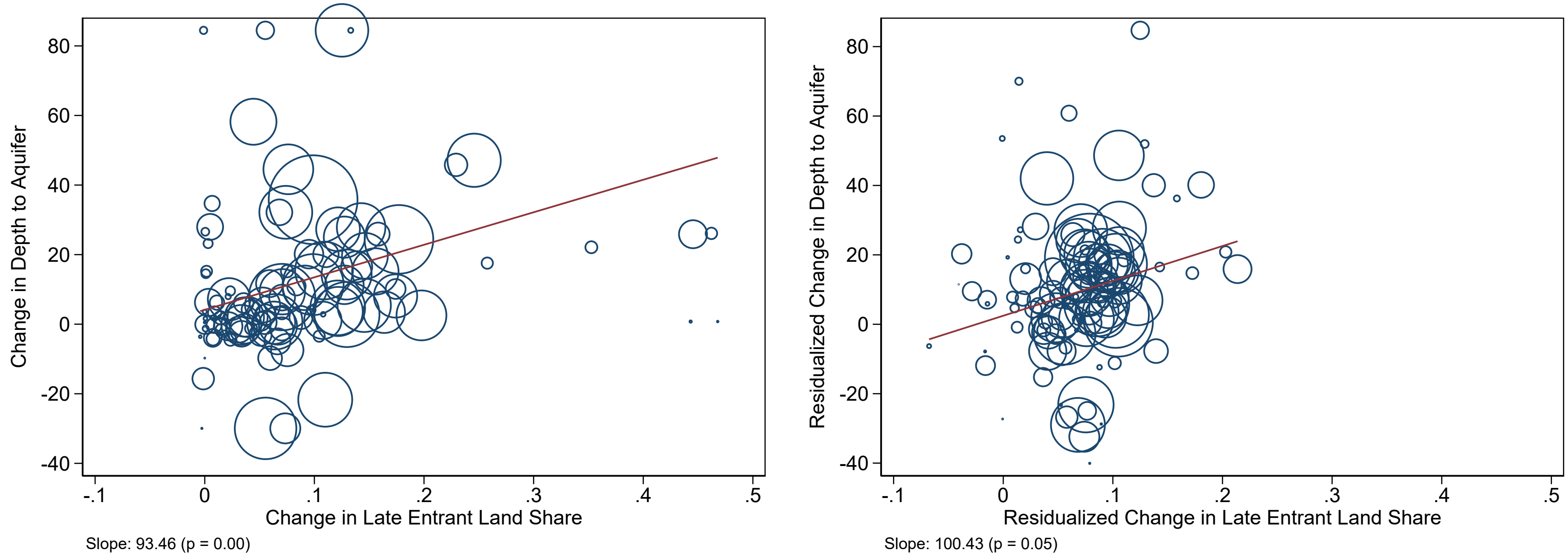


Note: The figure is analogous to Figure 4 except that it drops parcels smaller than 25 acres. Per the description of the name grouping algorithm in Appendix A.3, the algorithm is more likely to mis-classify individual, non-corporate growers, who plausibly own smaller plots of land, here measured as less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

Figure A.7: DAUCO-Level Relation Between Aquifer Depth and Late Entrants.

(a) Unconditional

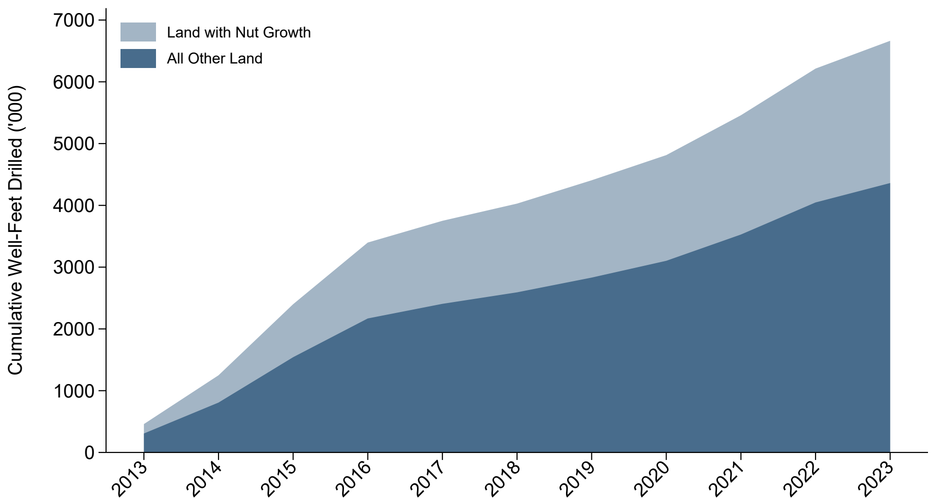
(b) Residualized



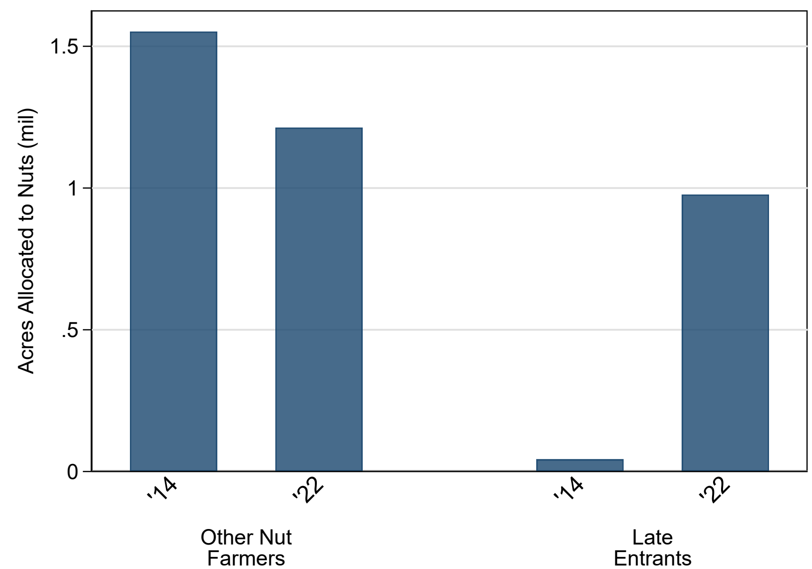
Note: This figure assesses robustness to how we construct measurement stations, which are our main unit of analysis, by plotting the relation between the 2013-2023 change in depth to the aquifer and expansion by late entrants to the nut market at the level of the detailed-analysis-unit-by-county (DAUCO). 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.8: 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
$SurfaceWaterRights_i$	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 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 4 and Appendix A.

Table A.3: 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.4: 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.5: 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.6: 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.7: 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.4(2)	5.15	5.22
... based on Table A.9(1)	5.26	2.83
Share of increase due to late entrants (η)		
... based on Table A.4(2)	16%	15%
... based on Table A.9(1)	14%	54%

Note: This table is analogous to Table 10, 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 (8) when using the estimates from Table A.9(1) to account for how they obtain from a log specification. The remaining notes are the same as in Table 10.

Table A.8: Using Nut Cropland Valuations to Forecast Returns.

	$TotalReturn_{n,t\rightarrow t+3}$		$IncomeReturn_{n,t\rightarrow t+3}$	
	(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.9: Miscellaneous Robustness.

	$\Delta\log(AquiferDepth_s)$	$\Delta AquiferDepthPrePeriod_s$			$TotalWater_{i,13\text{-}23}$
	(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 analogues of 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.