Domestic Well Failures and Bottled Water Demand

Katie Baker

December 2024

1 Motivation and Research Question

Public water systems provide residential and drinking water to most California residents, but many rural communities source their drinking water from private domestic wells instead. Between 3.4 and 5.8% (1.3 to 2.25 million) Californians rely on domestic wells to meet their water needs (Pace et al. [2023]). These wells are primarily located outside the bounds of public water infrastructure and are concentrated in low-income and Latino communities. Private wells are typically drilled to shallower depths than agricultural wells that draw groundwater from the same aquifers. This leaves domestic wells vulnerable to failing, or running dry, as groundwater tables decline. Hadachek et al. [2024] document that farmers in California extract more groundwater in response to heat and drought, which increases domestic well failures and imposes substantial externalities on low-income and minority rural communities. Dry private wells impose financial burdens on vulnerable households from the costly construction of new, deeper wells[1]. While a household is in the process of restoring its water supply, it faces additional costs to meet its water needs. Sometimes, local governments supply water or offer bottled water. Else, households must pay to truck in water, borrow water from a neighbor's well with a hose, or purchase bottled drinking water.

The risk of well failure will be exacerbated as groundwater tables continue to fall. The Sustainable Groundwater Management Act (SGMA), enacted in 2014, requires the development and implementation of groundwater sustainability plans to mitigate overdraft by 2040. However, analysis of the proposed plans in California's Central Valley aquifer system finds that nearly 10,000 private domestic wells and about 1,000 public supply wells would be impacted by declining groundwater levels allowable under the current proposals (Bostic et al. [2023]).

This paper will investigate the bottled water demand response to domestic well failures. We will match reported well failure reports with retail scanner data from nearby stores to quantify the impact of domestic well failures on bottled water purchases. Additionally, We will explore whether the price of bottled water products responds to well failure shocks.

2 Data

2.1 Domestic Well Data

Dry Well Reports

In 2014, the California Department of Water Resources (DWR) set up a system for households to report domestic well failures[2]. Owners usually only report issues when they cannot solve the problem on their own. These data contain the latitude and longitude coordinates for malfunctioning wells, the approximate date the issue started, and whether the issue was resolved. An upper bound on the duration of well outages can be inferred for resolved reports as the time between the issue start date and the report creation. For non-resolved cases, We will need to take a stance on the potential duration of the outage.

Nearly 6,000 domestic well failure reports were filed between 2014 and 2023. Reporting well failures is voluntary, so these data are an undercount of the true number of failed wells. While preparing the sample

[1] Domestic wells cost approximately $10,000 and are typically between 100 and 300 feet deep. Construction cost increases with well depth.

[2] Well failure reporting data is publicly available from 2014 through the present, accessible here

data, we drop failure reports missing the issue start date, with start dates that occur after the report was filed, and reports filed more than one year after the issue start date. Well failures that occurred but are not captured in our sample would bias our estimates downward.

Since domestic wells located near each other draw from the same groundwater basin, there is spatial correlation in well failures, particularly for wells drilled to similar depths. We plot the spatial and temporal correlation in well failure reports in section 5.

Well Completion Reports

The DWR also collects records for new well construction, repair, and destruction of wells. These data contain the coordinates of each well[3], the use (e.g., domestic water supply, agricultural irrigation), well depth, and when the work was completed. We will use these data to measure the total number of domestic wells in an area and the share of domestic wells that fail in a given week. We also plan to match the coordinates of well repair and construction reports to the dry well reporting data to measure the duration of shocks.

2.2 Scanner Data

There are two sources for water bottle purchase data we could use for the analysis. Both sources contain data from a subset of retail stores which are not representative of all stores in the area. Household purchases made at other stores will not be captured in either dataset.

Nielsen Retail Panel

The Nielsen Retail Scanner panel includes store-level weekly purchase data for bottled water and other beverage products from participating retail chains. The data include the price, number of units, and volume (ounces) of water sold at each store for weeks through the end of 2020. The store locations are identified by county and three-digit zip code. While most three-digit zip codes span multiple counties, the zip codes for the cities of Fresno and Bakersfield lie entirely inside Fresno and Kern counties. We will separate the stores in city zip codes in these counties from those outside of the city zip codes.

Using aggregated store data could lead to several potential issues. Lacking the location of stores in our sample, we cannot distinguish between stores near shocked wells and those further away. We also worry about potential spillovers. Consider a store located near a border where well shocks occur on the other side. It may be the case that a household travels across the border to the store in the control location.

Circana Data

We are pursuing a collaboration with colleagues at the USDA Economic Research Service that would allow us to access Circana weekly retail data from participating stores with longitude and latitude coordinates. With more precise store location information, we plan to match stores to nearby domestic wells within a specified distance.

Store-level Circana data would have several advantages over aggregated Nielsen data. With large geographic areas, we combine data from stores far from shocks into our response variable, which lowers the power of our tests. Although potentially sparse store coverage may limit the number of well shocks that fall close enough to a store to detect an effect, we will have better temporal coverage. Circana data are available through 2023, which includes the historic drought period in 2021 and 2022, when most of the reported failures in our sample occurred. Precise location data also addresses the potential issues arising from spillovers. We can determine the distance between the store and each dry well to compute the number (and share of total) wells that are malfunctioning within a reasonable driving distance. Circana data also include bulk shopping outlets and convenience stores which are likely to experience water bottle demand shocks.

Circana data will also allow us to identify stores that fall within public water system boundaries[4]. We plan to interact an indicator for the existence of public water infrastructure with the treatment to see how the effect varies with alternative water source availability. Finally, with our specification (discussed below),

[3]The DWR notes that most of the points have been spatially registered to the center of the 1x1 mile Public Land Survey System section that the well is located in.

[4]Public water system data come from California Drinking Water System Locations, a combination of three datasets that provides best-available location information for public California drinking water systems and state small drinking water systems. The data are combined into a GIS feature layer and can be accessed here

we will directly identify a store-level effect that is more interpretable than the aggregate effect over a larger geographic area.

2.3 Additional Datasets

Weather

We will control for weather using weekly drought severity maps from drought.gov and weekly average temperature maximums with data from PRISM. We plan to interact the drought indicator and temperature bins with the well failure index to explore how the treatment effect varies in different weather conditions.

Demographic Controls

We also obtain a set of demographic controls from the American Community Survey. These variables are available at both the county level and granular census block level. For three-digit zip code or location-specific Circana data, We will aggregate census block data to obtain estimates for the area of analysis. The variables We will include are population, median income per capita, median age, percent white non-Hispanic, percent Hispanic, unemployment rate, labor force participation rate, and percent of adults with at least college-level education.

3 Model

In the model, the weather (drought and heat, represented by **W**) impacts demand for bottled water (**Y**) directly and through the channel of increased groundwater extraction (**E**) for agriculture irrigation, which increases the likelihood of a well failure (**F**) that in turn raises demand. Demand is also influenced by several other factors (**X**), including the population size, income, and demographics near a given retail store, well water quality, and seasonality. We model bottled water demand as a function of price and acknowledge that prices may respond to realized or anticipated demand shocks. We plan to investigate the degree to which this happens in our analysis.

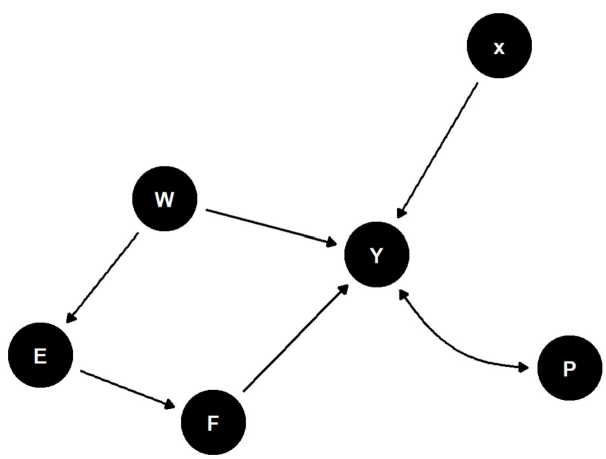


Figure 1: Directed Acyclic Graph of the Model

4 Estimation Strategy

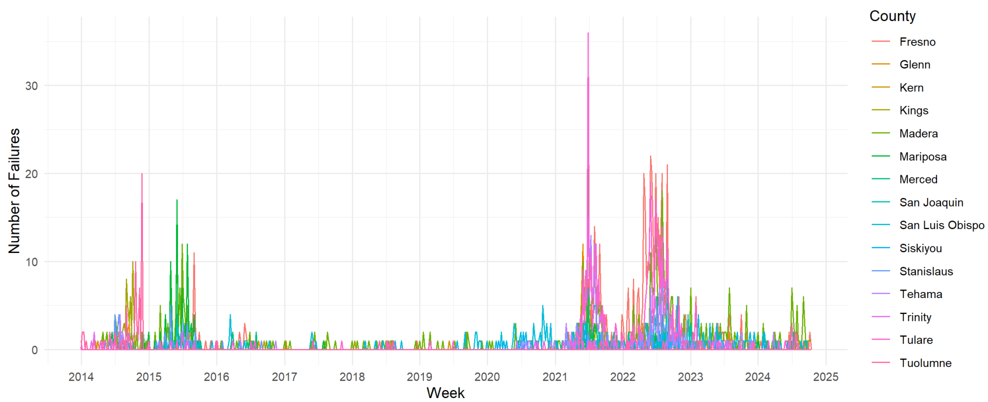
We plan to estimate the following specification:

$$Y_{l,w} = \alpha P_{l,w} + \beta_1 F_{l,w} + \delta W_{l,w} + \gamma X_l + \xi_l + \tau_w + error$$

The subscript l denotes the location (county, zip code area, or driving radius around store), and w indexes the observation week. $Y_{l,w}$ is the volume of bottled water in ounces purchased in location l in week w, divided by the number of stores in l. P is the average price per ounce. The well failure indicator F is the number or share of domestic wells experiencing failure each week and location. W includes a drought indicator and average weekly high temperature, while X includes the location-specific demographic estimates. We include fixed effects for location and week.

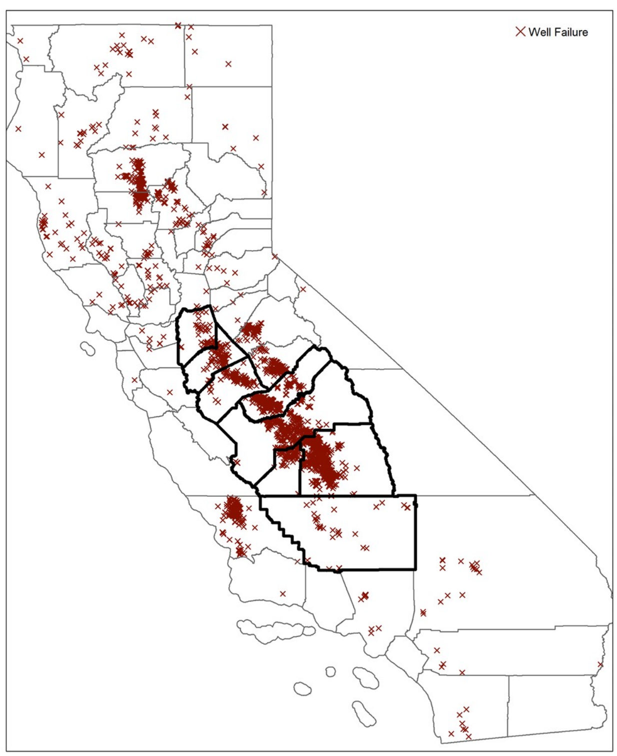
5 Spatial and Temporal Variation in Well Failure Reports

Figure 2: Weekly Well Failures by County, 2014-2024



Failures by week shown for the 15 counties with the most reported well failures in the sample.

Figure 3: Locations of Reported Well Failures, 2014-2020



Note: Counties in the San Joaquin Valley are demarcated with a thick border. Figure sourced from Hadachek et al. [2024]

References

Darcy Bostic, Linda Mendez-Barrientos, Rich Pauloo, Krisin Dobbin, and Victoria MacClements. Thousands of domestic and public supply wells face failure despite groundwater sustainability reform in california's central valley. *Scientific Reports*, 2023. doi: https://doi.org/10.1038/s41598-023-41379-9.

Jeffrey Hadachek, Ellen M. Bruno, Nick Hagerty, and Katrina Jessoe. External costs of climate adaptation: Groundwater depletion and drinking water. *Working Paper*, 2024.

Clare Pace, Carolina Balazs, Komal Bangia, Nicholas Depsky, Adriana Renteria, Rachel Morello-Frosch, and Lara Cushing. Inequities in drinking water quality among domestic well communities and community water systems. *Scientific Reports*, 2023. doi: https://doi.org/10.1038/s41598-023-41379-9.