The Impacts of Entry and Market Power in the California Retail Fuel Industry

Click Here for Most Recent Copy

Reid Taylor*

22nd November, 2023

Abstract

As California's gasoline prices consistently exceed the rest of the country, there has been significant interest among policymakers in understanding the role of market power in the state's retail gasoline market. In this paper, I estimate the effect of nearby entry and exit on the pricing of incumbent firms using high frequency price data and the precise geographic location of all gas stations in California. Using a difference-in-differences design, I find that entry of a new station is associated with a two-cent decrease in prices at incumbent stores, which equates to a 5% decrease in estimated retail markups. The effects are immediate, persistent, and show no sign of deterrence or limit pricing behavior. However, nearby exit results in precisely estimated null effects on prices. Both results are robust to various specifications and market size definitions. This paper also contributes to the conversation on California's "Mystery Gas Surcharge", a divergence between California and nationwide gas prices after the February 2015 Torrance refinery fire that cannot be explained by differences in taxes, regulation, or input and spot prices (Borenstein (2023)). I show that after prices did not respond differentially to the supply shock by market concentration and that the magnitude of entry effects were unchanged, providing evidence that retail competition was largely unaffected.

*Thank You!

*Department of Economics, University of California, Davis. rbtaylor@ucdavis.edu

1 Introduction

Understanding market power in the California gasoline market is of interest to policy makers and has been the subject of review by state agencies and the state department of justice for the past two decades due to prices that are consistently higher than the rest of the country.[1] Entry costs for new stations are high due to land values, environmental regulation, and local zoning laws. California blend rules result in an isolated market supplied nearly exclusively by the few in-state refineries. Further, these same refiners can have vertical relationships with downstream retail stores either through direct ownership of convenience stores or franchise contracts. Taken together, these factors suggest the potential for market power in oligopolistic local gasoline markets.

In this paper, I study competition in the California retail gasoline market by estimating the short-run effect of market size on incumbent firm pricing. Theory is ambiguous on the effect of market size and entry on prices Barron et al. (2004). Empirically estimating the impact of market size and firm entry on prices has traditionally been a challenge due to the endogeneity of market structure. Profit maximizing firms are attracted to markets with higher prices and profit margins. This selection bias results in cross-sectional studies yielding biased estimates for the causal effect of market size on price. Estimation has further been constrained by the availability of sufficiently rich data that contain numerous exogenous changes in market structure and granular price data Haucap et al. (2017). In light of price transparency regulations that have generated administrative data sets and the increased availability of third-party high-frequency data, modern panel data methods can be used to explicitly control for the inherent market structure.

To calculate the reduced-form causal effect of nearby competitors, I use historical daily station-level prices and precise geographic location for the universe of California gas stations from 2014-2018 with geographic and temporal variation in exposure to changes in the number of nearby competitors. The resulting data set includes over 700 new station entry and exit events and 14 million price observations. Using both difference-in-differences and event-study designs, I compare

[1] https://oag.ca.gov/antitrust/gasoline

the difference in prices at incumbent stations before and after a station enters or exits the market, with unaffected stations serving as the control group. I am able to control for the endogenous location decision of entering and exiting firms by including station, city-specific linear time trends, and day-of-sample fixed effects. The coefficient is thus identified by within-station variation in the number of nearby competitors and relies on the assumption that the exact timing of entry is exogenous conditional on these market characteristics.

I find that increased market competition reduces prices, but that the effect of changes in the number of competitors is heterogeneous. Entry of a new gas station nearby is associated with a two-cent reduction in gas prices at incumbent stations that is immediate upon the timing of entry. The amount represents an average of 5% of station markups during the sample period. Conversely, I estimate precise null effects of station exit on nearby incumbent station pricing. Event study results provide support for the key identifying assumption for causal interpretation, markets with and without entry or exit were following parallel trends in the periods leading up to the respective event.

I show that entry leads to a lower price equilibrium that is highly persistent in the market, lasting for years after the event, and present across all three blends of gasoline sold. Additionally, I show that the estimated effects are not attenuated by entry or exit of other stations nearby in the periods preceding, following, or simultaneous with entry and exit, lending credence to the market definitions. Estimated effects are highly localized and decay as the market definition broadens, consistent with prior literature on the nature of retail gasoline competition. There is no evidence that firms engaged in deterrence behavior or limit pricing in the periods leading up to entry.

Next, I present results focused on market power in the retail gasoline markets. Recent work has shown that gas prices in California rose sharply in response to the Torrance refinery fire and ensuing supply shock in February, 2015 (Borenstein (2023)). In the years since, an unexplained divergence between prices in California and the rest of the nation has emerged, averaging around 40 cents, resulting in $48 billion dollars of additional expenses for Californians as of 2023. This gap cannot be explained by differences in taxes or input prices and is concentrated after the point of wholesale distribution. I provide two pieces of evidence that suggest the supply shock had little effect on competition among retail stores and that market power of retail stations may not have been affected by the event. First, I provide evidence that the exogeneous shift in prices at the time of the Torrance fire was homogeneous across market concentrations. Secondly, I show that the estimated magnitude of entry effects for new stations was unaffected.

This work contributes to two strands of literature. First, this paper contributes to the growing literature on market size and entry in retail markets. Following Arcidiacono et al. (2020), similar studies using panel data methods emerged studying entry effects in gasoline markets focused on international markets where administrative price data was available. I add to this literature by presenting the first causal estimates of the effect market size and station entry on pricing in California. Results are similar in magnitude to other studies conducted in Mexico (Davis et al. (2023)) and Germany (Fischer et al. (2023)) also making use of high frequency, daily station data. Bernardo (2018) makes use of a liberalization of entry restrictions in Spain to study the impact of new entry on prices while González and Moral (2023) use an IV approach in the same setting.

Additionally, I contribute to the subset of the literature studying the unique California gasoline market. Stations in California, by regulation, must sell a special blend of gasoline (CARBOB), resulting in a distinct wholesale market from surrounding states. This has generated substantial literature documenting market power and pricing in the California market (Borenstein et al. (2004)), and vertical mergers (Hastings (2004); Taylor et al. (2010)).

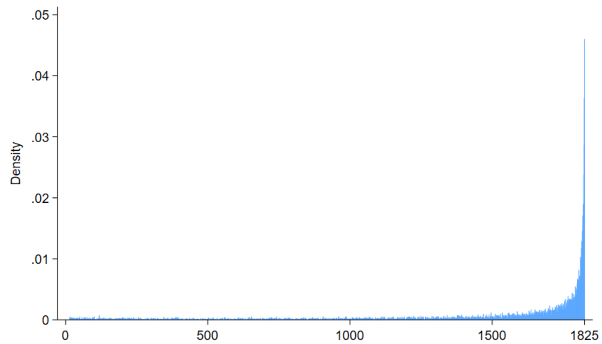
Understanding the effect of market structure in gasoline markets is also of importance to the conversation on the energy transition. As society moves towards reducing its reliance on fossil fuels, we need to understand how markets will be impacted. Supply-side restrictions on capital, either through reduced access to financial markets or through outright bans on fossil fuel extraction, transportation, and delivery have the effect of improving the position of legacy infrastructure. This is present in the California gasoline markets as several jurisdictions, starting with Petaluma in 2021, have banned the construction of new gas stations and restricted the expansion of existing stations. While gasoline demand is expected to decrease over time as electric vehicles market penetration increases, millions of gasoline-powered cars will still be operating on California's roads in the years to come and new gas stations are still in demand as shown by data. Banning entry of new stations results in counter-factually higher prices, increasing profits of incumbent firms.

In section 2, presents the data used in the empirical study. Section 3 discusses the empirical strategy. Section 4 presents the results of the estimate of entry and exit on incumbent pricing, as well as a series of robustness checks. Section 5 discusses the role of market power and Section 6 concludes.

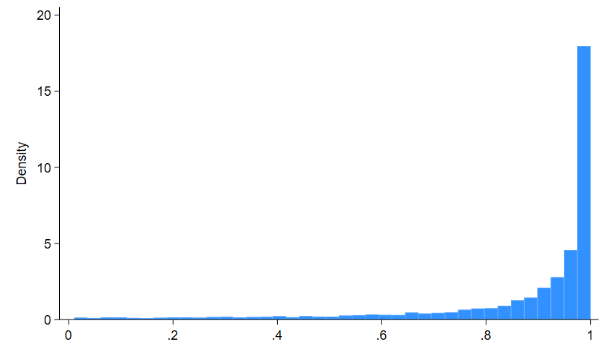
2 Data

To calculate the effect of market size changes on incumbent pricing, I use two main sources for data on gas stations in California. First, I obtain a panel of daily retail station gasoline prices in California from the Oil Price Information Service (OPIS) for years 2014 through 2018. I observe daily prices by fuel blend with the precise geographic coordinates for each station and a unique site identifier. Importantly, a station is identified in OPIS by its geographic location. Thus, the unique site identifier is unchanged when a store undergoes renovations, changes ownership, or changes store or fuel branding. This is important for the analysis, as entry or exit will not be confounded with changes that do not alter the number of competitors in the market. This however comes at the expense of being able to account for firm specific characteristics. For convenience, in the rest of the paper I use the terminology "station" to refer to the site where a gas station is located. To account for erroneous data, I exclude stations with less than 14 total price observations over the 4-year sample period as well as the 4 stations without geographic information. This results in a panel of 9,539 stations with over 14 million price observations.

Figure 1: Observations by Station: 2014-2018



(a) Observation Count



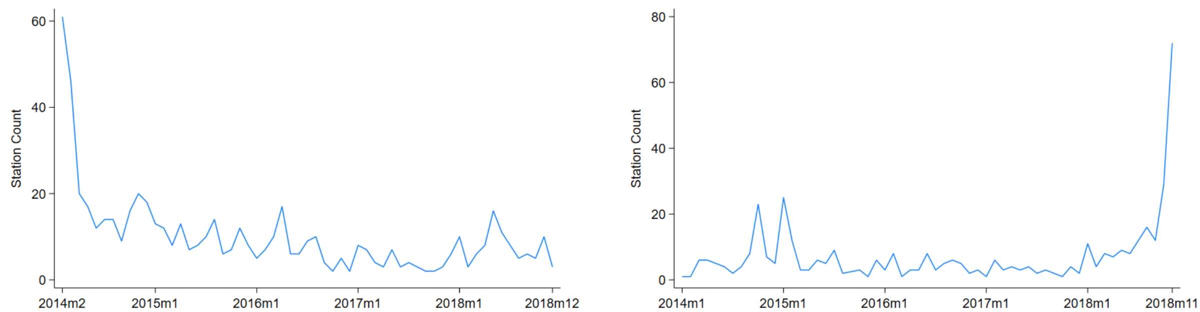
(b) % of Potential Observation

Notes: Data sourced from Oil Price Information Service (OPIS)

Prices in OPIS are collected through monitored fleet credit card transactions and direct feeds from stations. Only one price is reported per station per day per fuel type, however a price is not observed every day for each station or each fuel blend. OPIS reports prices for blends labeled as regular, mid-grade, and premium. As California cannot sell octanes below 87 or above 91, the terms directly map to a single octane blend. Separate from any attached convenience store branding, fuel can be branded or unbranded from the wholesale point. Since I do not observe the fuel branding of each station, estimated price effects at the time of entry or exit include price changes due to pure price competition and from potential contemporaneous blend changes.

In Panel (a) of Figure 1 shows the distribution of observation counts by station for the sample period 2014-2018 (1,825 days) for regular gasoline. Data coverage is high, with the median store having 1,756 prices reported. In fact, 75% of stores report more than 1,500 price observations over the sample period. To account for differential start and stop dates by station during the sample period, Panel (b) of Figure 1 reports the number of price observations as a percent of potential reporting days for each station using as a denominator the number of days spanned by the first and last observed price observations. The median store has prices reported for 97% of its respective sample period, and 75% of stores have prices reported for more than 86% of potential days.

Figure 2: Month of First and Last Observations by Station: 2014-2018



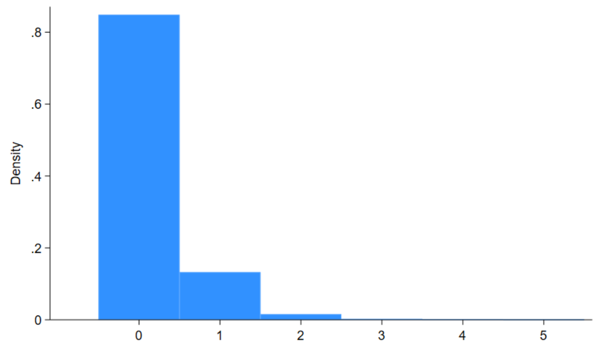
(a) First Observation

(b) Last Observation

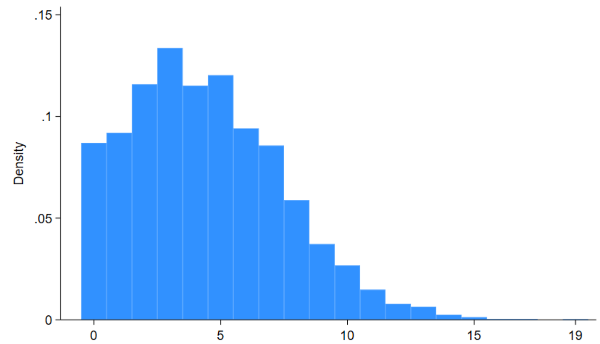
Notes: Data sourced from Oil Price Information Service (OPIS)

The OPIS data forms the basis for the empirical analysis on entry and exit. For each station, I calculate the entry date as the date of the earliest price observation and exit as the final date last price observation. Figure 2 reports firms by the date of their first and last price observation, grouped by months (January 2014 and December 2018 are excluded from the graph since the overwhelming amount of observation fall in these two months). There are a disproportionate number of entries and exits in the extreme tails of the sample period. This is likely driven by the fact that stations do not have prices reported every day and would have observations outside the window of the limited sample period. Therefore, I consider any station that has an initial price observation in January-March 2014 not to be an entry and any station to have their final price observation in November-December 2018 not to be an exit, and assume that they were present through the beginning or ending of the sample period. Additionally, stations that go dormant but have subsequent prices during the sample period are not considered as station exits or subsequent entry. This results in 484 unique stations having an entry date during the sample period and 348 unique station exits.

Figure 3: Market Characteristics by Station: 2014-2018



(a) Number of Entrants



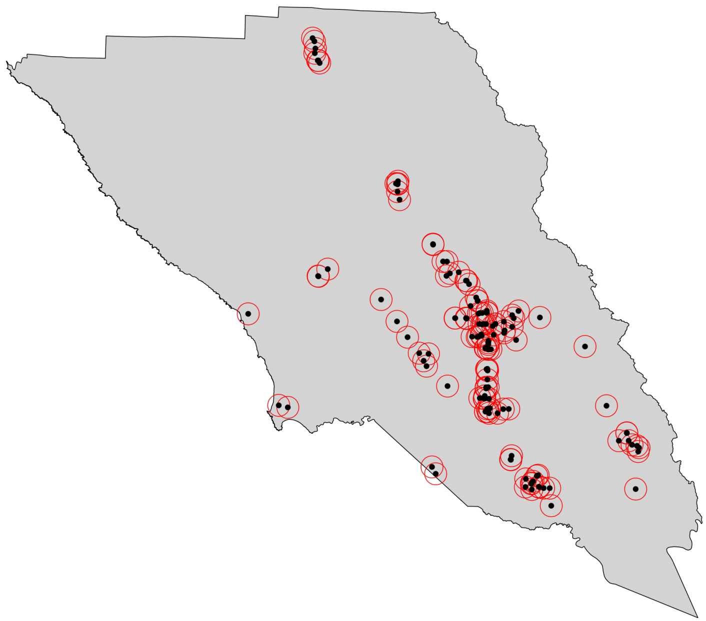
(b) Initial Market Size

Notes: Data sourced from Oil Price Information Service (OPIS)

Using the entry date, exit date, and geographic location of each station, I construct the main independent variable used in the empirical study, the number of competitor stations operating within a 1 mi. distance at the station-date level.[2] Panel (a) of Figure 3 shows the number of entrants experienced by stations. 85% of stations do not experience market entry during the sample period. Multiple entry is rare as 90% of stations that do experience entry only have 1 entrant during the sample period. As such, in the subsequent event study analysis, I focus on the first entry event observed during the sample period. Panel (b) of Figure 3 shows the distribution of initial market sizes for the full sample of stations. Markets contain relatively few stations, with the median store having four other gas stations within a 1-mile radius. There are stations which operate in monopoly markets and one station in downtown Los Angeles having 19 stations within a 1 mi. radius. 10% of stations are located in highly concentrated markets with 8+ stations. Appendix Figure A2 reports the cross tabulation of entrants by the market size. The majority of entries occur in markets with 3-8 competitors and are the only entrant during the sample period. Figure 4 below gives a visual representation of the variable construction, with gasoline stations in Sonoma county plotted in black with the 1 mi. market shown in red.

[2] Results for other distances measures varying from .5 to 10 miles are included as a robustness check.

Figure 4: Map of Gasoline Stations in Sonoma County: 2014-2018

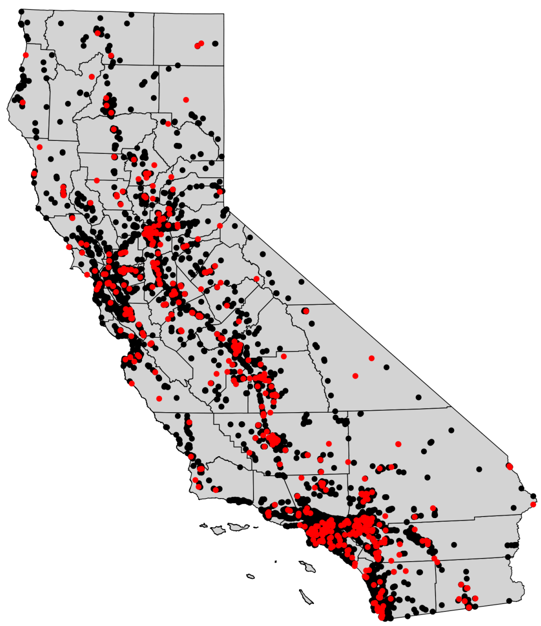


Notes: 1 mi. market definition shown in red on the map. Data sourced from Oil Price Information Service (OPIS)

The OPIS data report station and fuel branding, however the accuracy of the data is unreliable. Data from the Retail Fuel Outlet Annual Report (CEC-A15) conducted by the California Energy Commission include location, branding, operational time frame, store characteristics, and amenities. The Petroleum Industry Information Reporting Act (PIIRA) requires retail fuel locations to submit reports annually to the CEC. However, response rates are limited (60%), and thus are used in supplementary exploratory analysis.

There does not exist an official public source of gas station operation dates for the state of California and, to the best of our knowledge, the OPIS data represents the best source of information for the research presented. Collecting business permit data or underground tank information can locate stations but lacks the pertinent temporal component of on-site business operations which is likely the relevant metric of entry and exit for competing firms. State tax data on gas stations is reported at the owner later, and thus does not contain store-level information across multiple stores under the same ownership.

Figure 5: Map of Gasoline Station Entrants: 2014-2018



Notes: Entrant stations are shown by red dots, with incumbent stations represented by black dots. Data sourced from the Oil Price Information Service (OPIS).

There exists the potential for false positives in the OPIS entry and exit variable construction due to the limited window of the sample period and the fact that observations are when prices are reported, not official open or closing dates for the store. There also is the chance that a station goes unreported to OPIS. I validate the OPIS data by bench-marking the number of stations and change in stations over the sample period against two other measures for California: the Census Bureau County Business Patterns data and the estimated number of stations calculated by the California Energy Commission. The County Business Patterns data reports the number of businesses as of the week of March 12th of the appropriate year by NAICS code. The data show a net increase of 228 gas stations from 2014-2019. The CEC also undertakes an effort to estimate the number of stations in California based on returns from the A15 survey and other government data sources. For the same period, the CEC estimates a net increase of 190 stations in the state and around 10,000 gas stations. Neither source differentiates between entries or exits.

Figure 5 shows the location of the entrant gas stations in red and incumbent gas stations in black. Entry occurs in nearly all parts of the state, with concentration in the major urban areas, and along the I-5 and CA-99 corridors through the central valley. Visually, entry appears to largely match the locations of existing stations and not concentrated in new markets. There is observed entry in remote and rural parts of the state, however these events will only contribute to identification of the entry parameter in the empirical analysis if they are located sufficiently close to an incumbent station.

The main concern with previous cross-sectional analyses is that entry and exit are likely to occur in locations that differ from markets that do not observe a structural change. to test for baseline differences along observable demographic characteristics, I use demographic data from the 2014 American Community Survey. I compare census tracts that experience at least one entry or exit event to census tracts which have gas stations but experience no entry/exit during the sample period using the following cross sectional regression specification:

$$Char_t = \alpha + \beta Event_t + \varepsilon_t \tag{1}$$

which regresses the various demographics variables for tract t on a binary indicator variable for entry or exit in seperate regressions. I weight the regressions by tract population.

Table 1 reports the coefficients. Census tracts that experience entry and exit on average have lower household income, higher poverty rates, lower housing values, and more households than tracts with stations, but no events during the sample period. Tracts with entry are less likely to have zero-vehicle households, more likely to commute via vehicle, but have similar commuting times. Exit occurs in more populated tracts with lower income and housing values, more zero-vehicle households, and shorter commute times.

The baseline demographic differences between places that do and do not observe market composition changes highlight the need to account for the inherent market characteristics to address the endogenous entry and continuing operation decisions of the station. Additionally, to the extent that there are unobservable demographic characteristics that are correlated with both station entry and demand for gasoline, cross-sectional regressions of price on the number of competitors are

Table 1: Baseline Demographic Differences By Census Tract

	Entry	Exit
Income (Median)	-7,664.6***	-7,525.8***
	(1,359.8)	(1,644.3)
Income (Mean)	-11,109.9***	-7,702.9***
	(1,757.9)	(2,129.3)
Poverty Rate	2.874***	3.065***
	(0.548)	(0.664)
Households	325.9***	121.0**
	(45.8)	(55.6)
House Value (Median)	-111,373.0***	-50,330.9***
	(10,628.6)	(12,794.4)
% No Vehicle	-0.979***	0.911**
	(0.323)	(0.391)
% Commuting by Vehicle	1.290***	-0.148
	(0.485)	(0.587)
Commute Time	0.122	-1.660***
	(0.281)	(0.339)

Notes: Coefficients from cross-sectional regressions are shown, with standard errors below in parentheses. 2014 ACS demographic variables for census tracts are regressed on an indicator for whether the census tract experienced entry, in column 1, or exit, in column 2. Regressions are weighted by tract population.

likely to yield biased estimates of the effect of additional stations on prices due to the use of across station variation for identification.

3 Empirical Strategy

The localized nature of gasoline station competition allows for the geographic and temporal variation in exposure to station entry and exit across firms to form the basis of a difference-in-differences estimation for the causal effect on incumbent pricing. Following prior work by Arcidiacono et al. (2020), I treat the exact timing of the entry or exit of a new gasoline station as a short-run exogenous shift in the market structure for incumbent firms, conditional on the inherent market structure. In the preferred specifications, I define the relevant market as the 1 mile radius circle around the incumbent station. This market size is similar to prior literature which largely ranges between 1 and 2 miles (Lewis (2015); Davis et al. (2023); Fischer et al. (2023); Bernardo

(2018); Hastings (2004); Carranza et al. (2015); Barron et al. (2004)).

Importantly, I condition the model on a rich panel of fixed effects to account for unobserved variable bias inherent to the endogenous location decision of entering and exiting firms. By restricting the model to identifying variation within stations in the number of nearby competitors, over time, the model accounts for factors important to the locating decision such as the overall price level in the market, local price elasticity of demand, local traffic patterns, and relevant customer characteristics.

The estimating equation is:

$$P_{st} = \alpha + \beta N_{st} + \sigma_s + \delta_t + \Phi_c(t) + \varepsilon_{st} \tag{2}$$

where the main outcome variable is retail price at station s on day t and N_{st} is the count of stations within a 1-mile radius of station s on day t, increasing upon entry and decreasing with nearby exit. Separate regressions are run for the various blends of gasoline.

Station fixed effects (σ_s) are included to capture time-invariant differences between locations, such as station amenities and size, location effects, and distance to the wholesale terminal which largely drives differences in input costs. Day-of-sample fixed effects (δ_t) capture state-wide daily shocks to both input costs, such as oil-prices and refinery supply shocks, as well as daily shocks to product demand. Lastly, in the preferred specifications, I include a city specific linear time trend, $\Phi_c(t)$, which accounts for changes over times which may be correlated with price and station demand, flexibly across markets. Since the market definition varies from store to store, I cluster standard errors at the city level to account for common shocks across units.

The use use of unit and time fixed effects as in the above equation is known as the two-way fixed effects (TWFE) estimator. In the setting, stations that do not experience a change to market size within 1 mile during the sample period and previously treated stations both serve as control units for stations that experience entry or exit. Recent literature has shown that in the presence of heterogeneous treatment effects across treatment timing, the use of previously treated units as control units can lead to biased estimates. I return to this point later in Section 4.3 and show that results are largely unchanged in the robustness checks where I add increased flexibility to the TWFE equation as described in Wooldridge (2021) to account for potential heterogeneity. I report the main results using the TWFE specification from equation 2 for two main reasons: Highly localized nature of gas station competition makes homogeneous effects across treatment timing likely and the proposed solution estimators are computationally infeasible for the sample size with daily variation and requires significant aggregation.

Identification of the main coefficient of interest, β, as the causal effect of a change in the number of nearby stations on prices requires three main assumptions. First, the main identifying assumption requires entry and exit to be conditionally uncorrelated with the error term. Specifically, conditional on station fixed effects, time fixed effects, and time trends, the changes in station count are exogenous.

$$E[\varepsilon_s|\sigma_s, \delta_t, \Phi_c(t), N_{st}] = 0 \tag{3}$$

This also requires that there were no other factors perfectly correlated with the timing of the station entry that also impacted the pricing of nearby stations.

Secondly, the SUTVA assumption requires that there is no spillover of treatment onto control units outside of the impacted market. This is plausibly satisfied in this setting due to the local geographic nature of gas station competition, eliminating spillover price effects from treatment units to the larger pool of control observations. Lastly, identification in the differences-in-differences framework relies on the parallel trends assumptions. This requires that treatment and control units to be following common trends in the pre-treatment period and would have continued to do so in the absence of treatment.

In addition to the difference-in-differences model specified in equation 2, I estimate the following event study model for station entries, and equivalent analog for exit events separately:

$$P_{st} = \alpha + \sum_{k=-24}^{24} \beta_k \mathbb{1}[Entry_{st} = k] + \sigma_s + \delta_t + \Phi_c(t) + \varepsilon_{st} \tag{4}$$

setting event time indicators for the number of months before and after the first entry observed at station s. End points are binned to include 24 or more months before/after the event.

This additional specification has two attractive properties. First, it provides support for the parallel trends assumption needed for causal estimates. The concern being that if markets that observe entry or exit were not just different in terms of levels, but also changing over time differently, and these changes were correlated with the endogenous entry decision, this would introduce bias into the estimates. Precisely estimated null effects in the time periods leading up to entry or exit event provide support that treatment and control markets were following common trends. Secondly, I am able to decompose the effect of changes in station count by estimating separate regressions for entry and exit events.

The model specified above yields estimates of the average treatment on the treated (ATT) when the identification assumptions are satisfied. Given the baseline differences in treatment and control areas shown in Figure 1, we can think of the estimated coefficient as representing an internally valid estimate of the casual effect of entry or exit at locations where firms decide to enter or exit. The effect can be generalized to answer the question,"What will be the average effect of a station's planned entry?"

4 Results

4.1 Station Count Results

Results for the impact of changes in the station count within 1 mile on incumbent prices are shown in Table 2. Column 1 contains results from the simple regression of price on the station count variable as an initial point of comparison. There is an economically small, but significant negative relationship between market density and prices. The addition of a station within a 1 mi. radius lowers prices at incumbent firms by .6 cents. The specification fails to account for the endogenous relationship between price and entry, leading the estimate to likely be biased.

Columns 2-5 report results with added layers of fixed effects. In Column 2, the addition of a day-of-sample fixed effects to the naive regression accounts for daily common shocks to input

Table 2: Effect of Station Entry Within One Mile on Incumbent Pricing

Dependent Variable	(1)	(2)	(3)	(4)	(5)
Price	-0.006***	-0.006***	-0.067***	-0.012***	-0.010***
	(0.002)	(0.002)	(0.021)	(0.003)	(0.003)
R Sq.	0.001	0.809	0.163	0.956	0.958
Obs.	14,745,166	14,745,166	14,745,166	14,745,166	14,745,166
Log Price	-0.002***	-0.002**	-0.021***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.006)	(0.001)	(0.001)
R Sq.	0.001	0.802	0.165	0.951	0.953
Obs.	14,745,166	14,745,166	14,745,166	14,745,166	14,745,166
Station FE	No	No	Yes	Yes	Yes
Day of Sample FE	No	Yes	No	Yes	Yes
Linear Time Trend	No	No	No	No	Yes

Notes: OLS estimates for the effect of a change in the count of competitor stations within 1-mile on incumbent firm pricing for regular unleaded gasoline are shown. Data are from OPIS for years 2014-2018. Columns 1 reports estimates from a linear regression of the outcome on the station count. Columns 2-4 report results using the same model, but with the addition of the fixed effects listed in the panel below. Row one presents results where the outcome is price while row 2 presents results where the outcome variable is specified in log form. Model standard errors are reported in parentheses and clustered at the city level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

costs and demand such as crude oil prices or major weather events. While this accounts for a sizable portion of the variation in prices, as evidenced by the substantial increase in r-squared from Columns 1 to 2, the effect of entry largely remains unchanged. Column 3 accounts for inherent market characteristics that are endogenous to the firm's entry and exit decisions by restricting identification to within-station variation with the inclusion of the station fixed effects. An additional nearby competitor is now associated with a significant 6.7 cent decrease in prices. Columns 4 is the canonical two-way fixed effects model including both day-of-sample fixed effects and station fixed effects. Column 5 represents the preferred specification, showing results are robust to the inclusion of a city-specific linear time trend. An additional competitor within 1 mi. results in a 1 cent reduction in incumbent prices, which represents around a 2.5% reduction in firm markups during the sample period which average 40 cents.[3]

The evolution of coefficients across the table highlights the importance of addressing the endogenous entry decision of firms. Without the inclusion of station fixed effects, estimates rely on

[3]California Energy Commission estimates of CA gasoline price breakdown and margins.

cross-sectional variation in station counts. Once controlling for the inherent market characteristics with store fixed effect, chiefly whether entry occurs in a high or low-price market, the entry effect increases and is statistically distinguishable from zero. The change in the results with the inclusion of time effects highlights the need to account for seasonality in prices and entry timing as well as demand shocks. Lastly, a concern in this setting is that markets that are selected for entry or exit vary not only in terms of price levels, but are on separate growth trends, and that these trends are correlated with gasoline demand. The robustness of the coefficient to the inclusion of a city-specific linear time trend between Columns 4 and 5 provides support that treatment and control markets also do not vary in terms of trends.

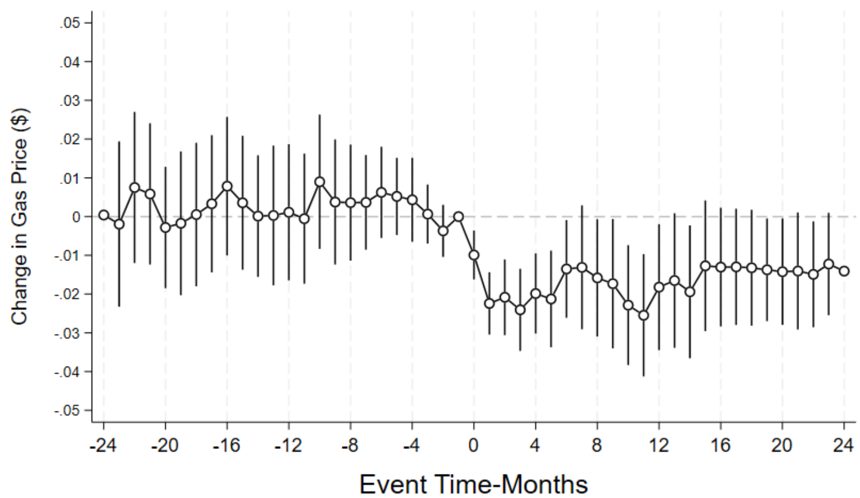
4.2 Event Study Results

We now turn to results from the event study specification shown in equation 4. Figure 6 presents results for the first entry event experienced by a given station, and Figure 7 reports results for the first exit event experienced during the sample period. An event study design is useful in this setting for multiple reasons. First, causal identification in a difference-in-differences model specification relies on the parallel trends assumption which requires treatment and control groups to evolve along common trends in the pre-treatment period and would have continued in the absence of treatment. The first part of this assumption can be tested visually and statistically via the event study design.

Point estimates for the event-time coefficients for the months prior to both entry and exit events are close to zero and all time periods include zero within the confidence intervals, supporting the identifying assumption. The null estimates in the pre-period also provide supporting evidence that firms do not engage in deterrence behavior or limit pricing when threatened by entry by dropping prices to make entry seem unappealing. While I do not have data on construction or permitting dates for new stations, the lack of any significant decrease in prices for years prior to arrival is likely sufficient. Firms only drop prices at the start of new operations and accommodate entry.

Secondly, by breaking out the entry and exit effect separately, we can see that the prior results from the difference-in-differences estimator, which is identified off of variation from both entry and exit events using the station count variable as the regressor, attenuated the effect due to heterogeneous effects by event type. Figure 7 shows precisely estimated null effects for exit events, which stand in contract to the negative effects of entry.

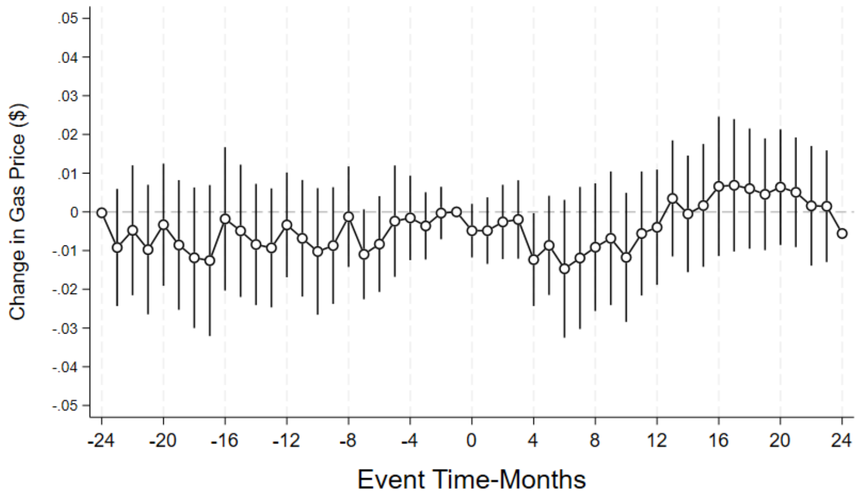
Figure 6: Effect of Station Entry on Incumbent Pricing



Notes: Coefficient estimates for the effect of station entry within 1 mile of an incumbent station on incumbent pricing of regular unleaded gasoline are shown. Event time t=-1 is the month prior to the arrival of the entrant. Coefficients are estimated from a linear regression of price on a panel of event time dummy variables, station fixed effects, day-of-sample fixed effects, and city-specific linear time trends. Standard errors are clustered at the city level.

Lastly, an event study design allows for estimates of the dynamics of the treatment over the duration of the post-period. Focusing on the entry effect in Figure 6, results show that entry of a station is associated with a sharp and immediate drop in the price of incumbent of around 2 cents. Effects are precisely estimated and persistent over the long run suggesting that the entry of a new station results in a quick shift to a new, lower price equilibrium.

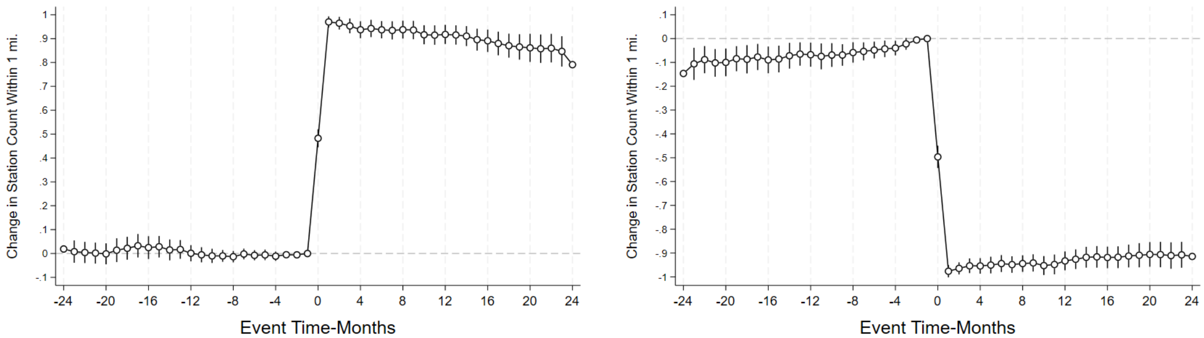
Figure 7: Effect of Station Exit on Incumbent Pricing



Notes: Coefficient estimates for the effect of station exit within 1 mile of an incumbent station on incumbent pricing of regular unleaded gasoline are shown. Event time t=-1 is the month prior to the departure event. Coefficients are estimated from a linear regression of price on a panel of event time dummy variables, day-of-sample fixed effects, and city-specific linear time trend Standard errors are clustered at the city level.

In both the difference-in-differences and event study models, results would be attenuated if the arrival of a new gas station were simply a replacement for a nearby exiting firm within the same market. In the extreme, perfectly timed entry and exit would not contribute to identification in the difference-in-differences model as the station count variable would remain constant across time. The event study design may also result in muted effects if entry and exit often accompany each other around the timing of the event. In spite of this, a negative relationship between entry and price could still exist in the presence of correlated entry and exit if the new station varies in product space with better site amenities or selling capacity compared to the exiting firm.

Figure 8: Effect of Station Entry and Exit on Market Size



(a) Entry (b) Exit

In Figure 8, I report the estimated coefficients from an event study specification similar to equation 4, but using the station count within 1 mile as the outcome variable, regressed on the event time dummies for entry events in Panel (a) and exit events in Panel (b) along with and station and day-of-sample fixed effects. Results suggest that there is no evidence of exits preceding entry events, shown by precisely estimated zero for event time coefficients during the pre-period. If there were significant exit preceding entry events, we would expect a downward sloping line above zero leading up to event time zero. At the time of the entry, the station count increases by 1 eliminating simultaneous entry and exit as a source of bias. The change in market size is persistent over the post-period with minimal decay in the store count up to one year post event. This rules out exit in the post-period as a driving force in the price results. Results for exit events are qualitatively similar, with slight evidence of leading entries in the market. At the time of the exit, there is no evidence of simultaneous entry, and no evidence of lagging entry in the year after the station exit event.

4.3 Robustness Checks

4.3.1 Two-Way Fixed Effect Advancements

In the presence of heterogeneous treatment effects and staggered treatment timing across units, recent work has shown that estimates of the average treatment on the treated from the canonical two-way fixed effects model, as outlined in equation 2, can be biased. This is due in part to the method's use of previously treated observations as control observations for later treated units, which is an incorrect counterfactual comparison in the presence of heterogeneous effects across treatment cohorts. While this is not likely to be a concern in my setting since the majority of observations come from never-treated stations, I confirm that results are robust to newer estimators below.

Many of the proposed alternative estimators rely on estimating treatment effects for each treatment cohort immediately before and after treatment non-parametrically to ensure appropriate control units are used in estimation. These approaches are computationally inefficient in the presence of many treated cohorts, which are defined by the exact date of entry in the current data specification. Additionally, to the extent that there is any mismeasurement of the exact date of entry or exit, these methods can yield biased estimates. Collapsing the data to higher levels of aggregation eases the computational constraint as cohorts can be defined by common month or quarter of entry, however this strips the model of the essential identifying variation of daily prices.

I follow Davis et al. (2023) in using the solution proposed in Wooldridge (2021) to show that results are robust to increased levels of flexibility to the TWFE model. The method fully saturates the TWFE model by estimating coefficients for treatment-cohort and time interaction terms in addition to the unit and time effects from the TWFE specification, parametrically estimating the heterogeneous treatment paths across cohorts. These group coefficients can then be aggregated across treatment cohorts to calculate the average treatment effect of the treated parameter.

Focusing first on the effect of entry on price in Row 1 of Table 3 below, Column 1 reports the coefficient from a version of the TWFE regression in equation 2 using a treatment indicator variable for the timing of the first entry event with station and quarter-of-sample fixed effects as a point of direct comparison for the later Wooldridge (2021) estimator. Entry is associated with a 2.2 cent decrease in incumbent pricing, similar to the results shown earlier from the event study. In Column 2, I show results are robust to the correction used in Davis et al. (2023) which increases flexibility in the model by including city-specific linear time trends and a treatment cohort-specific linear time trend, where treatment cohorts are defined by common quarter-of-sample for the entry event.

Table 3: Effect of Station Entry Within One Mile on Incumbent Pricing

Dependent Variable	(1)	(2)	(3)	(4)
Entry	-0.022***	-0.019***	-0.024***	-0.027***
	(0.004)	(0.004)	(0.003)	(0.004)
Exit	0.002	.002	-.001	-.001
	(0.004)	(0.004)	(0.003)	(.004)

Notes: Estimated coefficients on dummy variables for the first entry and exit experienced by incumbent firms are estiamted in separate regression and shown by row. Column 1 reports the baseline TWFE estimates. Column 2 adds city-specific linear time trends and treatment cohort time trends. Column 3 implements the estimator from Wooldridge (2021) designating treatement cohort at the quarter level, while Column 4 uses a monthly treatment cohort designation, but only a 5% sample of never-treated station.

Next, in Columns 3 and 4 I report the estimated average treatment effect of the treated parameters from implementation of the full estimator from Wooldridge (2021). Designation of treatment cohorts below the quarter-of-entry level was computationally infeasible using the full sample of over 14 million observations. Column 3 reports the aggregated ATET using the quarter-of-entry to designate treatment cohorts. In Column 4, I report the aggregated ATT using month-of-entry to designate treatment cohorts using the full sample of treatment observations, but a restricted 5% random sample of never treated stations for control observations. Results obtained from the TWFE specifications are robust to the correction methods, and similar results are produced from both levels of aggregation in the Wooldridge (2021) estimator.

Turning now towards the exit effects in the next row of Table 3, the estimated null effects from the TWFE model are also robust to both the use of the Wooldridge (2021) estimator and the level of cohort designation in Columns 3 and 4.

4.3.2 Other Blends

To this point results have focused on the most commonly sold blend of gasoline, regular unleaded 87 octane gasoline, which represents around 70% of motor gasoline sales in California. In table 4, I present results for the other two main blends of gasoline, mid-grade (89 octane) and premium (91 octane). Results closely mirror the outcomes for regular blend gasoline shown above. Nearby entry is associated with a 1.6 cent price decrease for mid-grade gasoline, while premium gasoline decreases by 1.5 cents. as with regular gasoline, exit events are not associated with statistically distinguishable changes to incumbent pricing.

Table 4: Effect of Entry and Exit by Fuel Blend

	Regular	Mid-Grade	Premium
Count	-0.010***	-0.009***	-0.008**
	(0.003)	(0.003)	(0.003)
Entry	-0.020***	-0.016***	-0.015***
	(0.004)	(0.004)	(0.004)
Exit	0.000	0.003	0.001
	(0.004)	(0.004)	(0.004)
Station FE	Yes	Yes	Yes
Day of Sample FE	Yes	Yes	Yes
Linear Time Trend	Yes	Yes	Yes

Notes: Coefficients for the station count variable, the first entry and the exit experienced by station are estimated separately and shown by row. Columns show the blend of gasoline used for the outcome variable. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

4.3.3 Distance

Turning focus back to the regular grade gasoline, I report results from varying the distance measure used to define the market size surrounding a station in Table 5. Column titles show the length of the radius, in miles, used to define the circular market. Results evolve across the distances as expected with the strongest effects calculated for the narrowest market definition. Results suggest that effects quickly dissipate with increasing market size.

Table 5: Effect of Entry and Exit by Distance

	.5	1	2	3	4	5	10
Count	-0.011***	-0.010***	-0.009***	-0.008***	-0.006***	-0.006***	-0.005***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Entry	-0.021***	-0.020***	-0.018***	-0.019***	-0.013***	-0.008*	-0.011*
	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)
Exit	0.002	0.000	0.001	-0.001	0.000	0.001	0.006
	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)

Notes: Effects of station count, entry, and exit on incumbent prices are reported for various market definitions. Column titles represent the size of the circular market definition around each station in miles. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

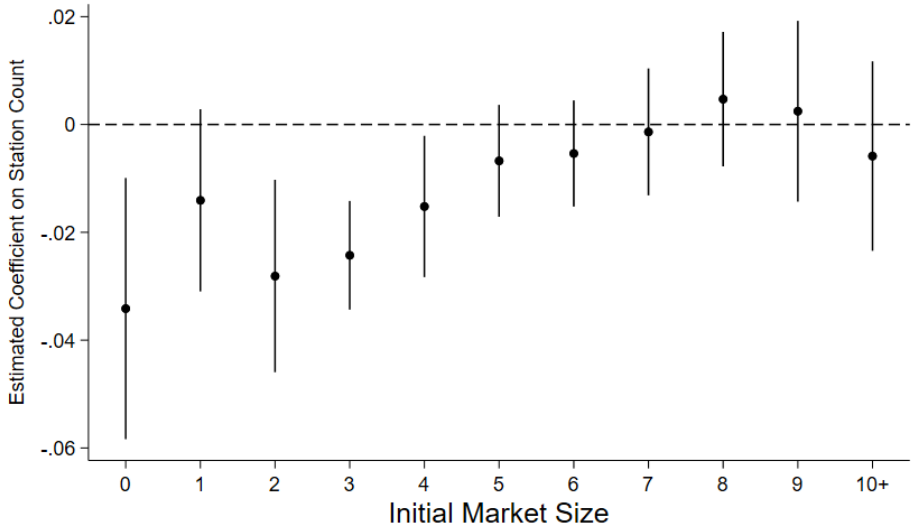
5 Discussion

5.1 Market Concentration

Canonical work by Bresnahan and Reiss (1991) on competition in homogeneous goods markets suggests that entry into smaller, consolidated markets results in larger competitive effects and that this effect dissipates as the number of competitors in a market increases. However, Barron et al. (2004) shows that this result can reverse in search model with price dispersion depending on the search costs involved. I test for heterogeneous effects of entry across market sizes in my setting by estimating Equation 2 separately by initial market size.

Results using the preferred specification of day-of-sample, station fixed effects, and city specific time trends are shown in Figure 9. Effects dissipate towards zero, mostly monotonically, as market density increases. Effects are concentrated among markets with less than 5 competitors within a 1-mile radius. This in fact aligns with the findings in Bresnahan and Reiss (1991) which found entry above 5 firms led to no changes the competitive conduct of incumbent firms in their setting. Given the lack of data on station characteristics, I am unable to claim if the price effect is due to eroding market power or a result of heterogeneous station characteristics.

Figure 9: Effect of Station Entry Within 1 Mi. on Incumbent Pricing



Notes: Coefficient estimates for the effect of station entry within 1 mile of an incumbent station on incumbent pricing of regular unleaded gasoline are shown. Coefficients are estimate by running separate regressions by initial market size. Regressions include day-of-sample, station fixed effects, and city specific time trends. Standard errors are clustered at the city level.

5.2 Mystery Surcharge

California has the highest gasoline prices in the country, and research has recently focused on trying to explain the divergence in prices between the state and the rest of the country. The price of crude oil, traded on a global market, is common across the country, as is the federal gas tax, and thus do not contribute to the difference in prices. Much of the difference is explained by California regulation. California's gasoline excise tax, sales tax and environmental fees from cap-and-trade, low carbon fuel standards, and underground storage tank fees combine to be 2-3 times greater than the average state-level taxes in the rest of the country. Lastly, California has strict emission standards on gasoline sold within the state, resulting in a blend (CARBOB) that is uniquely used in California. The blend, almost uniquely supplied by refiners located in California, results in refinery level prices higher than in the rest of the country. However, even after decomposing the components of the pump-level gasoline price and accounting for the observable differences, there remains a portion of the price difference that is unexplained.

Recent work by Severin Borenstein has noted that prior to the Torrance refinery fire on February 18, 2015, the difference in pump-level prices between California and the rest of the country was nearly exclusively explained by the components noted above. After the fire, the ensuing supply shock resulted in prices increasing in California. However, prices never converged back in line with national prices once the refinery resumed activity. An unexplained price difference averaging 40 cents per gallon emerged, a phenomenon referred to as the ”Mystery Gasoline Surcharge”. This has resulted in $48 billion dollars of additional expenditures on gasoline for Californians as of 2023. The mystery surcharge is concentrated in the supply chain after the gasoline has left the refinery or wholesale rack, which leaves the retailing industry, contracting between refiners and retailers, and transportation costs between wholesale and the pump as potential sources of the difference.

If station markups increase, this can potentially impact the entry effect of a new station. Using the timing of the refinery fire as an exogenous supply shock, we can modify the difference-in-differences equation above to test for differential entry effects before and after the refinery fire by including an an additional term interacting the post-entry indicator with a post-fire indicator. The time fixed effects absorb the other standard terms required in a difference-in-differences setting.

$$P_{sct} = \alpha + \beta_1 Entry_{sct} + \beta_2 Entry_{sct} * Post_t + \sigma_s + \delta_t + \rho_c(t) + \varepsilon_{sct} \quad (5)$$

The outcome variable, P_{st} is price at the station-date level. The coefficient, β_1 reflects the reduced form effect of the entry of a station within a 1-mi. radius of station s. The coefficient β_2 represents the incremental change to the entry effect experienced by firms post-fire, with $\beta_1 + \beta_2$ capturing the full effect of entry in the post period. As before, it is important to include station fixed effects to account for the endogeneity of the location of entry.

Results are presented in Table 6 below. In column one, I present the baseline estimation of the entry effect using the difference-in-differences analog to the event study results shown above for entry. Entry within a 1 mi. radius is associated with a 2.0 cent decrease in incumbent pricing on average over the full sample period. Column 2 reports the estimated coefficients from estimation of equation 5. We fail to reject similar magnitudes of entry for events which occurred after the refinery fire, shown by the coefficient on Entry x Post being small and statistically insignificant.

Table 6: Effect of Station Entry Post Torrance Refinery Fire

	(1)	(2)
Entry	-0.020***	-0.024***
	(0.004)	(0.009)
Entry x Post		0.005
		(0.010)
Station FE	Yes	Yes
Day FE	Yes	Yes
Linear Time Trend	Yes	Yes
R Sq.	0.958	0.958
Obs	14,745,166	14,745,166

Notes: OLS estimates for the effect of entry within 1-mile on incumbent firm pricing for regular unleaded gasoline are shown. Data are from OPIS for years 2014-2018. Column 2 interacts an indicator for post Torrance refinery fire with the post-treatment variable. Model standard errors are reported in parentheses and clustered at the city level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

6 Conclusion

Using daily price data and the timing of the entry of new gas stations and exit of existing gas stations, I estimate the effect of market size changes on the pricing for incumbent stations. The use of high-frequency data and the ability to restrict identifying variation to within station allows the approach to account for the endogenous entry and exit decisions of firms. I find that an increase in market size from entry is associated with statistically significant 2 cent decrease in the price at incumbent stations. This is compared to a precise null effect for the exit of a nearby firm. Both results are robust to various specifications, new estimators that correct for heterogeneous treatment effect in the two-way fixed effects specification, and across the various blends of gasoline sold. I also show that the results are strongest for the closest entries and dissipate as the market definition broadens. Exit is a precise null effect at all market definitions. These results are in line and of similar magnitude to studies in other countries (Davis et al. (2023); Fischer et al. (2023)).

I also contribute to the conversation surrounding market power in the California gasoline market. California has gas prices that have diverged from the rest of the country for unexplained reasons after the Torrance refinery fire in Februrary 2015. I provide some of the first evidence in the literature discussing changes in market conditions after this exogenous shock by showing that the magnitude of entry effects are unchanged. If the unexplained markup were concentrated at the retail pump, then we may expect larger entry effect in the post-period. I additionally contribute to the discussion by showing that prices did not change differentially across market sizes after the fire, further bringing into question if retailers are the source of the unexplained markup.

References

Arcidiacono, Peter, Paul B Ellickson, Carl F Mela, and John D Singleton, "The Competitive Effects of Entry," *American Economic Journal. Applied Economics*, 2020, *12* (3), 175–206.

Barron, John M., Beck A. Taylor, and John R. Umbeck, "Number of sellers, average prices, and price dispersion," *International Journal of Industrial Organization*, nov 2004, *22* (8-9), 1041–1066.

Bernardo, Valeria, "The effect of entry restrictions on price: evidence from the retail gasoline market," *Journal of Regulatory Economics*, feb 2018, *53* (1), 75–99.

Borenstein, Severin, "What's the Matter with California's Gasoline Prices?," 2023.

_ , James Bushnell, and Matthew Lewis, "Market Power in California's Gasoline Market," *CSEM Working Paper*, 2004, *132*.

Bresnahan, Timothy F and Peter C Reiss, "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 1991, *99* (5), 977–1009.

Carranza, Juan Esteban, Robert Clark, and Jean François Houde, "Price controls and market structure: Evidence from gasoline retail markets," *Journal of Industrial Economics*, 2015, *63* (1), 152–198.

Davis, Lucas W, Shaun McRae, and Enrique Seira, "The Competitive Effects of Entry in the Deregulated Mexican Gasoline Market," *Working Paper*, 2023.

Fischer, Kai, Simon Martin, and Philipp Schmidt-Dengler, "The Heterogeneous Effects of Entry on Prices," 2023.

González, Xulia and María J. Moral, "Competition and Competitors: Evidence from the Retail Fuel Market," *The Energy Journal*, sep 2023, *44* (01).

Hastings, Justine S., "Vertical relationships and competition in retail gasoline markets: Empirical evidence from contract changes in Southern California," *American Economic Review*, mar 2004, *94* (1), 317–328.

Haucap, Justus, Ulrich Heimeshoff, and Manuel Siekmann, "Fuel prices and station heterogeneity on retail gasoline markets," 2017.

Lewis, Matthew S, "Odd Prices at Retail Gasoline Stations: Focal Point Pricing and Tacit Collusion," *Journal of Economics and Management Strategy*, sep 2015, *24* (3), 664–685.

Taylor, Christopher T., Nicholas M. Kreisle, and Paul R. Zimmerman, "Vertical relationships and competition in retail gasoline markets: Empirical evidence from contract changes in southern california: Comment," jun 2010.

Wooldridge, Jeffrey M., "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators," *SSRN Electronic Journal*, aug 2021.

A Appendix: Figures

Appendix Figure A1: Month of Last Price Observation by Station

		Number of Entrants						
		0	1	2	3	4	5	Total
Initial Market Size	0	783	43	3	1	0	0	830
	1	781	81	12	3	0	0	877
	2	968	124	9	3	0	0	1,104
	3	1,084	163	25	2	0	0	1,274
	4	919	158	19	1	1	0	1,098
	5	970	153	15	2	6	1	1,147
	6	765	120	9	2	1	0	897
	7	636	165	14	2	0	0	817
	8	438	101	13	7	2	0	561
	9	289	54	12	0	0	0	355
	10	207	37	11	0	0	0	255
	11	114	25	2	0	0	0	141
	12	57	16	1	1	0	0	75
	13	48	13	0	0	0	0	61
	14	16	6	1	1	0	0	24
	15	9	3	0	0	0	0	12
	16	3	1	0	0	0	0	4
	17	3	0	0	0	0	0	3
	19	1	0	0	0	0	0	1
	Total	8,091	1,263	146	25	10	1	9,536

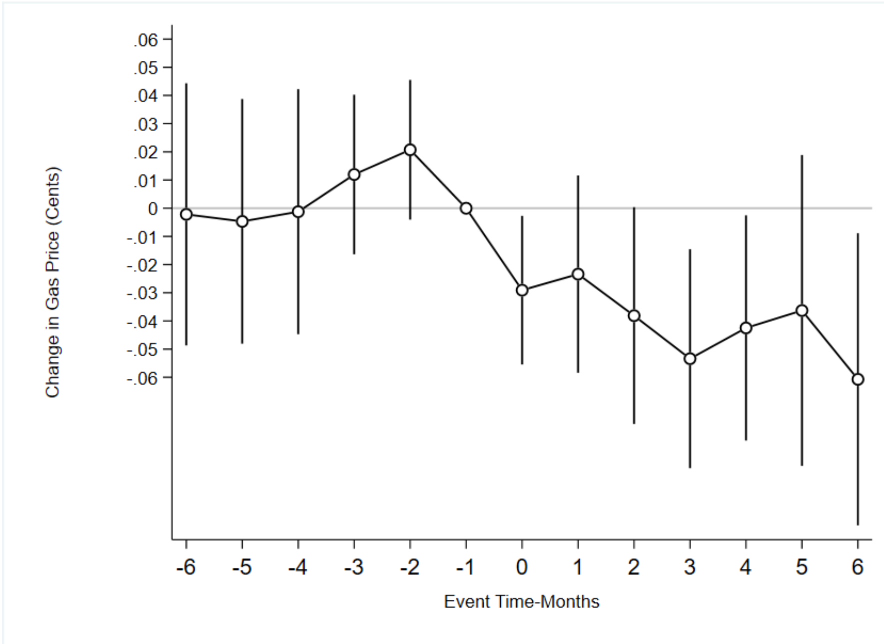
Notes: The number of entrants experienced for each station is shown by the initial market size for the incumbent station.

Appendix Figure A2: Month of Last Price Observation by Station

		Number of Entrants						
		0	1	2	3	4	5	Total
Initial Market Size	0	783	43	3	1	0	0	830
	1	781	81	12	3	0	0	877
	2	968	124	9	3	0	0	1,104
	3	1,084	163	25	2	0	0	1,274
	4	919	158	19	1	1	0	1,098
	5	970	153	15	2	6	1	1,147
	6	765	120	9	2	1	0	897
	7	636	165	14	2	0	0	817
	8	438	101	13	7	2	0	561
	9	289	54	12	0	0	0	355
	10	207	37	11	0	0	0	255
	11	114	25	2	0	0	0	141
	12	57	16	1	1	0	0	75
	13	48	13	0	0	0	0	61
	14	16	6	1	1	0	0	24
	15	9	3	0	0	0	0	12
	16	3	1	0	0	0	0	4
	17	3	0	0	0	0	0	3
	19	1	0	0	0	0	0	1
	Total	8,091	1,263	146	25	10	1	9,536

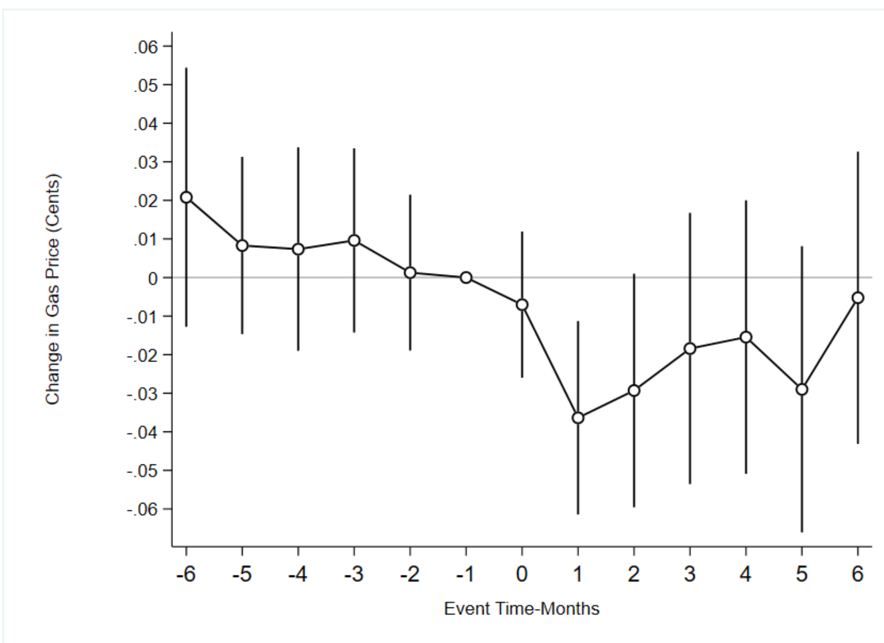
Notes: The number of entrants experienced for each station is shown by the initial market size for the incumbent station.

Appendix Figure A3: Event Study for Market Size 0



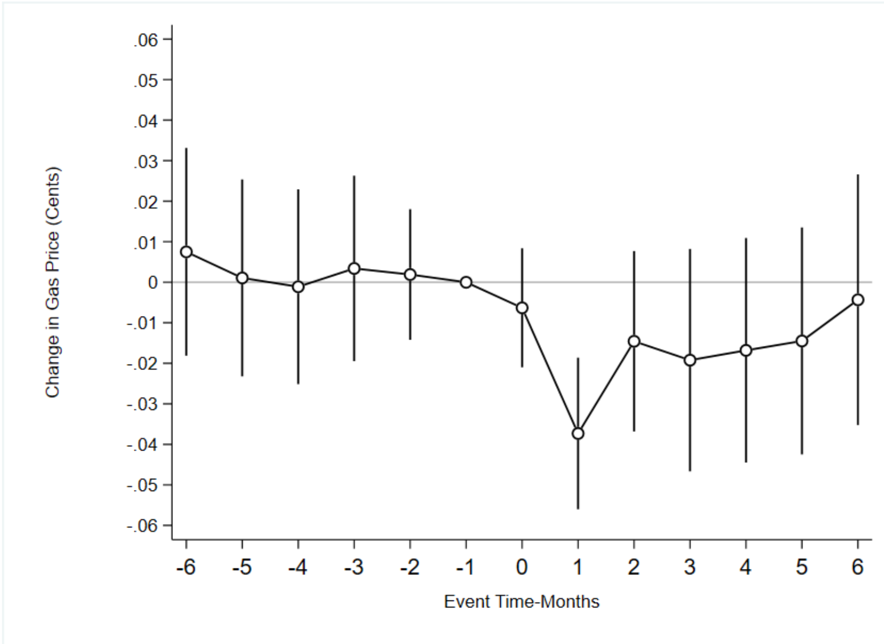
Notes: Event Study for the effect of firm entry on incumbent's pricing with a initial market size of 0.

Appendix Figure A4: Event Study for Market Size 1



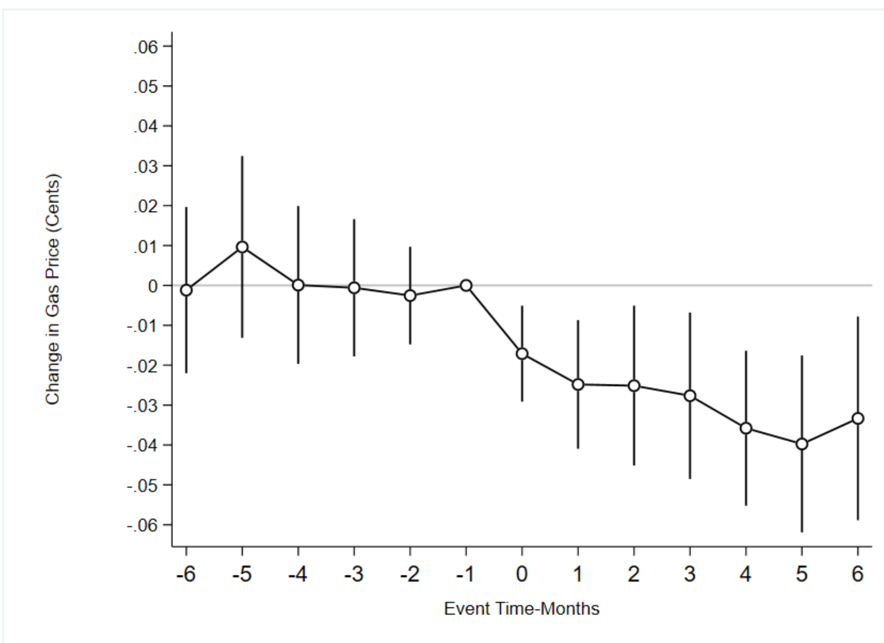
Notes: Event Study for the effect of firm entry on incumbent's pricing with a initial market size of 1.

Appendix Figure A5: Event Study for Market Size 2



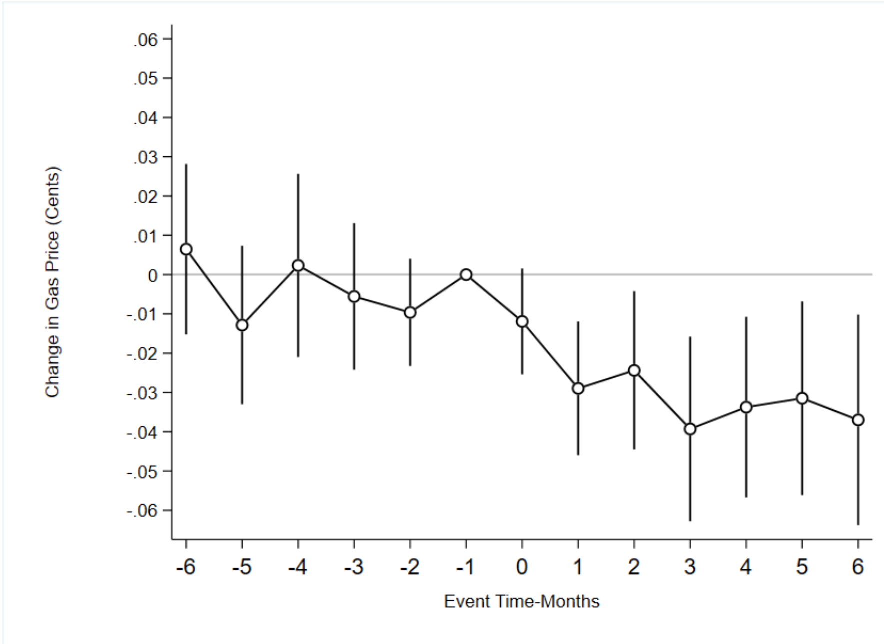
Notes: Event Study for the effect of firm entry on incumbent's pricing with a initial market size of 2.

Appendix Figure A6: Event Study for Market Size 3



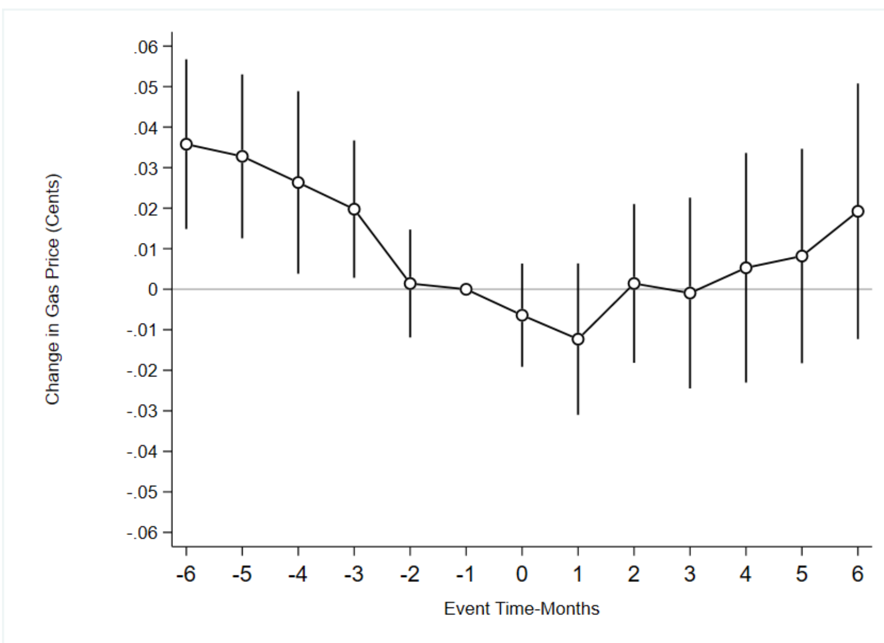
Notes: Event Study for the effect of firm entry on incumbent's pricing with a initial market size of 3.

Appendix Figure A7: Event Study for Market Size 4



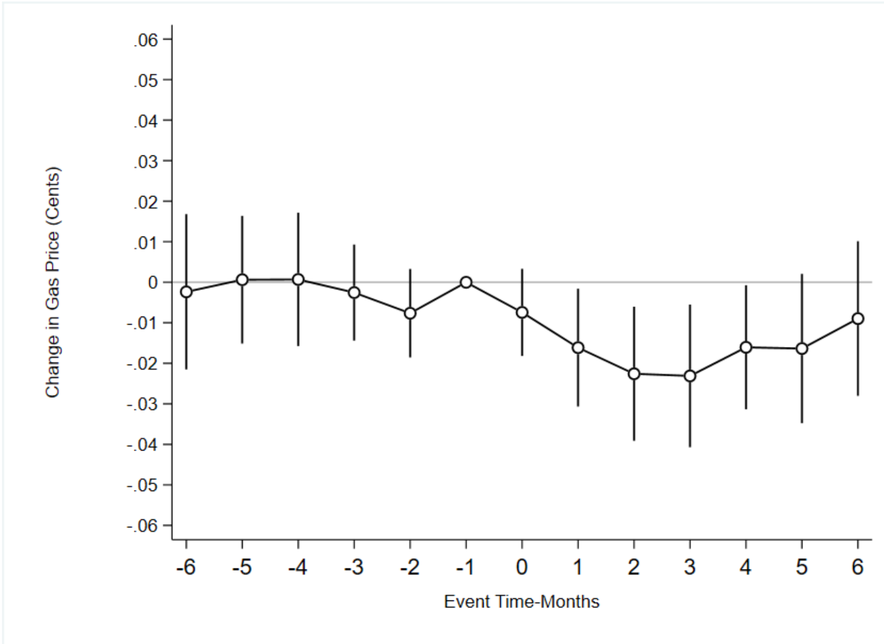
Notes: Event Study for the effect of firm entry on incumbent's pricing with a initial market size of 4.

Appendix Figure A8: Event Study for Market Size 5



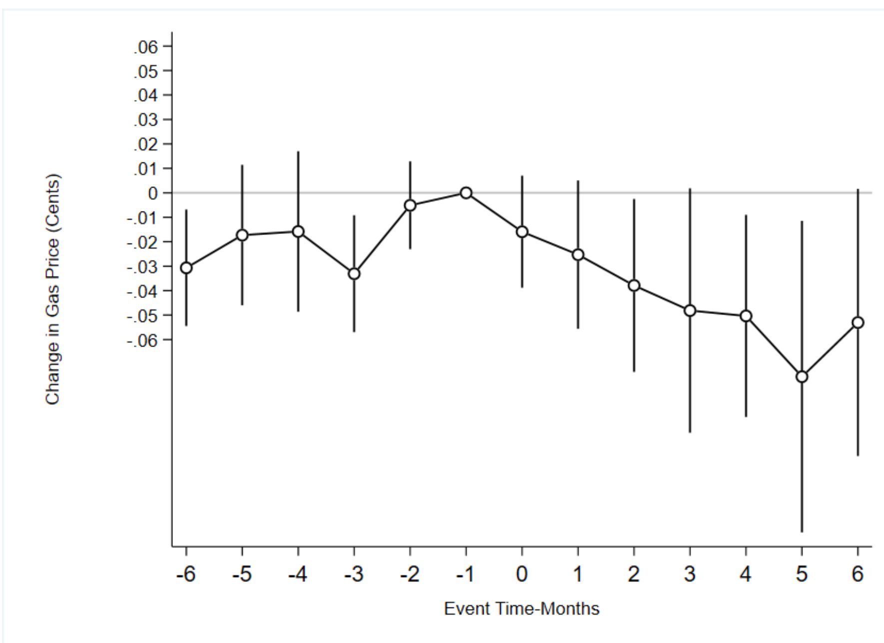
Notes: Event Study for the effect of firm entry on incumbent's pricing with a initial market size of 5.

Appendix Figure A9: Event Study for Market Size 6-10



Notes: Event Study for the effect of firm entry on incumbent's pricing with a initial market size of 6-10.

Appendix Figure A10: Event Study for Market Size 11+



Notes: Event Study for the effect of firm entry on incumbent's pricing with a initial market size of 11+.