The Dynamics of Competition in the Gasoline Market

Nicholas Stafford

Master's Thesis

Professor Colman

12 May 2025

Abstract

In this paper, I aim to estimate and clarify the effect of local competition in the market for gasoline. In particular, I am interested in testing the validity and applicability of the law of competition in the context of your local gasoline market. Researchers Eckert and West concluded that gas stations in local markets collude to set prices contrary to the competitive market model (235). After collecting price and location data from gas stations on way.com, this paper examines the effect of the geographical layout of a local market on the array of prices that market charges. At the 1% level of significance, two separate methods of measuring competition had a reliable and consistent impact on gas prices. The first measure of competition, the number of stations in a quarter mile, half mile, and single mile radius, decreases prices by 4.78, 4.57, and 2.12 cents per gallon, respectively, for every additional station. The second measure, the distance to the next three closest stations increases prices by 2.37, 9.04, and 7.58 cents per gallon for each mile of separation from closest to third closest respectively. The last measure, the difference in prices between competing stations, is a function of distance between the stations. The expected increase in the spread is 4.99, 2.60, and 5.29 cents per gallon for each extra mile separating the closest three stations respectively holding quality of gasoline and city constant.

Introduction

Neoclassical economics suggests that as the number of firms increases in a market, the price of the good or service decreases because consumers have more options and therefore more leverage. In other words, firms that sell the same product compete for the same consumers. In this paper, I assume a gallon of regular unleaded gasoline is identical at all gas stations. In a market where several firms sell the same product with little to no differentiation, each firm must offer a price better than its competitors if it wants to sell its product to consumers. Conversely, the greater the distance between competitors, the less leverage the consumer has because their choice adjusts to show the added marginal cost of that distance. This idea is central to the theory of spatial economics. The added cost of geography whether temporal, financial, laborious, or otherwise is reflected in prices. In this case, the prices of gasoline increase when the consumer's opportunity cost decreases because of the greater distance.

This paper asks the question, "What is the effect of competition between gas stations in close proximity on the price they charge for a gallon of regular unleaded gasoline?" My hypotheses are:

1. The price of gasoline will decrease as the competition in the neighborhood increases, and

2. As the consumer's decision between competing firms becomes easier, i.e. as the cost of travelling between stations decreases, the observed prices will decrease as well.

3. Stations located far away will have less of an effect on/association with the observed station compared to stations nearby.

Using the OLS estimation method and panel data, I will illustrate the effects of local competition on gas prices at 928 stations in 10 cities across the United States. The paper will follow with a brief literature review to discuss previous work in the field, a data overview and

description of the models used, a section for regression results and interpretations, and finally a conclusion.

Literature Review

In 2005, Eckert and West studied the impact of several variables on the odds that a gas station in Vancouver would match the mode price. One of their variables of interest was the number of competing stations within two kilometers, which they hypothesized would have no impact on the likelihood of that station matching the mode price. Similarly, I calculated a variable for the number of competing stations within a half mile; however, I anticipated the coefficient to reflect a decrease in price as the number of firms increases. This reasoning follows from the idea that number of firms in the market is a positive factor of the supply curve. Thus, as both number of firms and subsequently supply increases, the equilibrium price will decrease. The same is thought to be true for the local markets in this study.

In order to understand the reasoning behind my selection of distance to the next closest gas station, one must consider the characteristics of spatial competition. Spatial competition is based in the theory of opportunity costs. The consumer is balancing the cost and benefit of going to the next best choice of gas station. Capozza and Van Order phrased the phenomenon well: "If transportation were costless, firms would have no protection from spatially separated rivals. A firm 1,000 miles away would be just as formidable a competitor as one next door" (897). In our world, however, transportation poses many, often large costs to consumers and firms alike. This cost plays a significant role in the dynamics of product pricing by gas stations. I aim to estimate the effect of a minimal transportation costs faced by the consumer to reach a competing station.

Pinkse et al. investigated the dynamics of the gasoline market on both large and small scales. One of their conclusions was that while there are domino effects that lead stations far

apart to indirectly influence each other's prices, markets are generally extremely small spatially (1141). This paper will assume this understanding when examining the effects of competition. In addition to the main variables, I recorded the distance to the farthest in-city competitor for each station and their price difference. My hypothesis is in line with Pinkse's conclusion that the price differential should be greater for distant competitors than for local competitors.

A recent paper by Taylor and Muehlegger sought to understand the effects of entry and exit in the retail gasoline market with a difference-in-differences analysis. Using panel data from 2014-2018 inclusively, they looked at the post treatment effect of station entries and exits, and they propose that entries into a local market reduce prices by 2.5 cents per gallon and exits have a negligible effect on prices. The latter conclusion is unusual. Taylor and Muehlegger argue that the conclusion could be a result of branding differences between entries and exits or even that those stations exiting exerted less competitive pressure on the market before they shutdown. The former conclusion, on the other hand, is aligned with the law of supply and demand. The number of firms is part of the function of supply. As the number of firms increases, supply increases, and an increase in supply results in lower prices and higher quantity sold. While this study does not have the data required to examine entries and exits exactly, it will attempt to explain the same phenomenon through a cross-sectional analysis of number of stations in the local market as well as the pressure competitors put on each other as a function of distance.

Data

This study considers the prices and locations of 928 stations for 193 days across ten midsized cities. The data was webscraped from way.com, a one-stop shop for all resources that a car owner may need to save money. In this case, I used the site's fuel price reporting service specifically. Those mid-sized cities are:

- Atlanta, GA
- Bakersfield, CA
- Cleveland, OH
- Colorado Springs, CO
- Fresno, CA
- Milwaukee, WI
- Orlando, FL
- Raleigh, NC
- Sacramento, CA
- Wichita, KS

The dataset has 175,704 observations after accounting for duplicates provided by way.com. Of the observations, 53,340 (30.36%) do not have reported prices. The dependent variable for the first two hypotheses, price, is measured in cents per gallon of the stations' regular unleaded gasoline. The dependent variables for the other hypothesis are the difference in prices in cents per gallon between each station and its first, second, and third closest competitors (diff_first, diff_second, and diff_third). Additionally, this paper includes an analysis of the effect of distance ('farthest') on the price differential (diff_far) between each station and its farthest in-city competitor to further understand if a domino effect exists as Pinkse et al. suggest.

Of the independent variables that affect price directly, the first are three counts of each competitor in the neighborhood/local market. The variables, named radius_quarter, radius_half, and radius_whole, count the competing stations within a radius of a quarter mile, half mile, and one mile. Each radius excludes the smaller range(s) it encircles if applicable. In other words, the one mile range is the number of stations at most one mile from the observation excluding those counted in the half and quarter mile ranges already. The same is true for the half mile radius which excludes those stations within a quarter mile. In the end, the three ranges avoid double counting and specifically act as bands or zones around each station. The goal of looking at all three radii is to examine the effect of local competition while defining how large the local market is.

The second independent variable set of interest allows for distance variation by considering only the distance in miles to the closest, second closest, and third closest competing stations (closest_first, closest_second, and closest_third) within the list of stations. Similarly, the last set considers the differences between prices of neighboring stations and the correlation with closest_first, closest_second, and closest_third. One notable change is the analysis of stations far apart as a comparison case to those close together.

To start, I'll list the state and city controls: population (pop), measured in thousands, is provided at the city level as of 2020 by the Census Bureau as well as land area which was used to calculate population density (density). Density, measured as persons/sq. mile, is simply the original population divided by the city's land area. State PCE and price parity were given by the Bureau of Economic Activity for the year 2023. State PCE is otherwise known as the per capita personal consumption expenditure price index in thousands, and I have included it in this study to account for personal wealth of consumers. Price parity is an estimation of each state's cost of living as a proportion of the US average. For example, with the US average being 100 as it is in this study and Ohio being 91.8, what would cost a consumer on average $100 will only cost them about $91.80 in Ohio. The West Texas Intermediate (WTI) crude oil index is a global index of crude oil measured in USD per barrel, and the one-month futures contract for WTI was collected daily from bloomberg.com. Concluding the state and city controls is the gas tax which was provided by the American Petroleum Institute as of 1/1/2022 and incorporates all state taxes and fees on non-diesel gasoline. A drawback to many of these controls is that the cities included in this study may not exhibit the same or similar traits as the state-level data would imply. Finally, I used a control for quality of gas sold called "top_tier". In the gasoline industry, certain brands are authorized and compelled to follow fuel additive standards that cause less carbon build up in

the automobile's engine track. This standard is set by Top Tier™ and more info can be found at toptiergas.com. BP and Amoco stations left the Top Tier ecosystem in 2013, but they claim to still uphold and even exceed Top Tier standards with their Invigorate® detergent formula. After careful evaluation, I included the BP and Amoco brands in the "top_tier" category.

Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
price	122,364	3.370	0.719	1.870	5.690
diff_first	92,192	0.183	0.240	0.000	2.310
diff_second	90,229	0.205	0.247	0.000	2.310
diff_third	86,005	0.198	0.240	0.000	2.410
diff_far	84,762	0.248	0.251	0.000	2.110
radius_quarter	175,704	0.622	0.765	0	5
radius_half	175,704	0.912	1.042	0	6
radius_whole	175,704	3.319	2.590	0	15
closest_first	175,704	0.348	0.315	0.0002	1.982
closest_second	175,704	0.594	0.344	0.038	2.316
closest_third	175,704	0.783	0.382	0.104	3.098
farthest	175,704	7.683	1.521	4.423	12.412
top_tier	175,704	0.627	0.484	0	1

Notes: price and price differences are in USD/gallon; differences in absolute values; distances are in miles

There are three OLS-estimated regression models in this paper. The third follows the same structure but is modified to examine the prices differences and distances of the other two closest competitors as well as those of the farthest competitor. The models are as follows with "A" representing the set of state and city controls:

I. $price = \beta_1 radius\_quarter_s + \beta_2 radius\_half_s + \beta_3 radius\_whole_s + A + u_s$

II. $price = \beta_1 closest\_first_s + \beta_2 closest\_second_s + \beta_3 closest\_third_s + A + u_s$

III. $diff\_first = \beta_1 closest\_first_s + \beta_2 top\_tier_s + \beta_3 city_s + u_s$

Each of the first two models is included as a pooled OLS and a panel regression with entity fixed effects by station to account for the variation in price due to time-varying measures

that affect every station simultaneously. The only difference being that the pooled OLS regressions also include another beta for the WTI. The first model focuses on the number of stations in the local market, while the second focuses on the distance to the nearest three alternative gas stations. The intent behind the first model is to illustrate that the number of options the consumer has in the immediate area will impact the prices they observe. The second model tries to illustrate the effect of the choice between a few stations near each other.

After running a Breusch-Pagan test for heteroskedasticity on both models, it is apparent that each is suffering from heteroskedasticity with both showing p-values of <0.001. Another issue with the data is the autocorrelation in the prices. The Durbin-Watson test also exhibited p-values of <0.001 in both cases. In an effort to correct these problems, I implemented heteroskedasticity-consistent standard errors in all four regressions using the field-standard Arellano method. This method accounts for both heteroskedasticity and entity autocorrelation. Lastly, Hausman tests for efficient and consistent coefficients suggest that the models are both efficient with the information available and consistent between random effects and entity fixed effects. The tests had p-values between 0.75 and 0.80.

The third model was inspired by the consideration that in a perfectly competitive market where geography is a negligible factor in a consumer's choice, the consumer will choose the product going for the lowest price. This is not to say that geography is not a factor; however, in a scenario where two stations are in the immediate vicinity and, therefore, the consumer faces negligible increased costs to either choice, we would expect that consumer to purchase from the firm charging the lowest price. The model investigates the effect of distance to each of the first three competitors on any station's relative pricing behavior. In a theoretical world, the limit of the differences in price should be 0 as the distance approaches 0. In other words, one station

competing with another in the same location selling the same product should also charge the same price for that product. This exact relationship is my hypothesis for the third model. Specifically, the hypothesis is that each additional mile of separation between two stations will increase the difference in their prices. If the hypothesis cannot be trusted, the most likely culprit would be that consumers perceive the quality of gasoline to be better or worse for certain stations or brands – that the market does not behave as perfect competition.

Results

Once again, the first two hypotheses are that the number of gas stations in a neighborhood is negatively correlated with gas prices and that the distances to the closest competing gas stations are positively correlated with gas prices. Both regressions for the first model signify that the first hypothesis is true, while those for the second model also signify that the second hypothesis is true.

Local Market and Distance Regression Results

	Dependent variable:			
	price			
	POLS 1	FE 1	POLS 2	FE 2
radius_quarter	-4.875***	-4.775***		
	(0.115)	(0.113)		
radius_half	-4.554***	-4.567***		
	(0.084)	(0.056)		
radius_whole	-2.125***	-2.116***		
	(0.037)	(0.026)		
closest_first			2.491***	2.373***
			(0.384)	(0.205)
closest_second			8.748***	9.036***
			(0.512)	(0.299)
closest_third			7.875***	7.576***
			(0.419)	(0.207)
gas_tax	2.516***	2.498***	2.487***	2.470***
	(0.017)	(0.036)	(0.017)	(0.035)
wti	-1.401***		-1.393***	
	(0.018)		(0.018)	
pop	0.108***	0.107***	0.102***	0.101***
	(0.001)	(0.004)	(0.001)	(0.004)
density	-0.194*	-0.162	-0.322***	-0.303
	(0.113)	(0.191)	(0.113)	(0.191)
state_pce	0.064	0.007	0.072	0.016
	(0.045)	(0.089)	(0.046)	(0.089)
price_parity	2.643***	2.720***	2.705***	2.779***
	(0.056)	(0.077)	(0.057)	(0.076)
Constant	18.307***		-10.400***	
	(2.843)		(2.928)	
Station FE	No	Yes	No	Yes
Observations	122,364	122,364	122,364	122,364
R^2	0.830	0.837	0.826	0.833
Adjusted R^2	0.830	0.837	0.826	0.833
F Statistic	66,373.440*** (df = 9; 122354)	78,558.730*** (df = 8; 122163)	64,348.550*** (df = 9; 122354)	76,050.260*** (df = 8; 122163)

Note: *p<0.1; **p<0.05; ***p<0.01
Robust standard errors in parentheses.

The coefficient on radius_quarter suggests that for each additional station within a quarter mile the station's price decreases by 4.78 cents per gallon holding state and city controls constant. Similarly, each additional station in the half mile band, or the next quarter mile, decreases prices by 4.57 cents per gallon holding state and city controls constant. The effect of a station in the second band has a remarkably similar magnitude to that of the first band. Given

how close the coefficients are for each band, a logical conclusion would be that the local market reasonably extends to at least a half mile in every direction. Even in broad terms, the model suggests that a station is less affected by those competitors farther away because those additional stations in the more distant zones have a smaller magnitude of effect on prices.

On the other hand, the expected effect of each additional mile between the observed station and its closest competitor is an increase of 2.37 cents per gallon holding state and city controls constant. Unexpectedly, the magnitude of the coefficients for the second and third closest stations -- 9.04 and 7.58 cents per gallon respectively -- are greater than for the closest station. While the direction of the effect agrees with my hypothesis, I had expected the closest competitor to impact prices more than the second or third closest competitors.

The reasoning is as follows: in a local market, the consumer has a choice between any number of firms to buy the good they desire. Let us assume a market with four gas stations each selling the same gallon of regular unleaded gasoline. This resembles a perfect competition model where the utility-maximizing consumer will go to that firm which is selling the product at the lowest price, and firms must charge accordingly or risk losing business. As previously mentioned, we must consider geography as a factor because it adds an extra dimension to the choice. Intuitively, the farther the station, the more gas and time -- among other things -- the consumer must consider in their opportunity cost. Is the longer trip worth the savings in gasoline? If a consumer is at station A, the regression model suggests that an extra mile to the second closest station C plays a larger role in the price charged than an extra mile to the closest alternative B. This result seems to indicate that having a second competitor in the local market is more important than the distance to the nearest. The absence of a third and fourth firm increases the price per gallon by much more than the absence of a second firm.

Below I have listed the regression results for the price differential models:

Price Differential Regression Results

	Dependent variable:			
	diff_first	diff_second	diff_third	diff_far
	1st Competitor	2nd Comp.	3rd Comp.	Farthest Comp.
closest_first	4.988***			
	(0.179)			
closest_second		2.603***		
		(0.250)		
closest_third			5.291***	
			(0.238)	
farthest				-0.022
				(0.071)
top_tier	1.363***	1.317***	1.198***	0.440**
	(0.109)	(0.151)	(0.104)	(0.210)
factor(city)bakersfield-ca	18.201***	22.088***	17.917***	22.079***
	(0.356)	(0.432)	(0.429)	(0.479)
factor(city)cleveland-oh	0.110	-0.377	-0.449	-9.174***
	(0.302)	(0.397)	(0.417)	(0.348)
factor(city)colorado-springs-co	7.208***	1.724***	1.242***	-2.775***
	(0.379)	(0.348)	(0.405)	(0.342)
factor(city)fresno-ca	26.466***	21.702***	23.901***	19.997***
	(0.873)	(0.653)	(0.572)	(0.834)
factor(city)milwaukee-wi	-5.174***	-4.866***	-6.142***	0.818
	(0.433)	(0.512)	(0.410)	(1.754)
factor(city)orlando-fl	-2.648***	-3.426***	-5.377***	-12.010***
	(0.408)	(0.421)	(0.485)	(0.306)
factor(city)raleigh-nc	-4.610***	-7.241***	-5.900***	-8.509***
	(0.323)	(0.329)	(0.379)	(0.422)
factor(city)sacramento-ca	20.337***	20.744***	17.687***	6.053***
	(0.425)	(0.484)	(0.395)	(0.366)
factor(city)wichita-ks	-9.680***	-10.389***	-10.748***	-11.917***
	(0.300)	(0.318)	(0.362)	(0.350)
Station FE	Yes	Yes	Yes	Yes
Observations	92,192	90,229	86,005	84,762
R^2	0.233	0.241	0.230	0.238
Adjusted R^2	0.231	0.240	0.228	0.237
F Statistic	2,534.032*** (df = 11; 91988)	2,602.943*** (df = 11; 90025)	2,328.783*** (df = 11; 85801)	2,406.943*** (df = 11; 84558)

Note: *$p<0.1$; **$p<0.05$; ***$p<0.01$

Robust standard errors in parentheses. City reference is Atlanta, GA; differences are in cents/gallon

Note that the number of observations in these models is less than the previous models because these can only calculate a difference when both stations show a price for the day.

For each additional mile between the observed station and its nearest competitor, we can expect the difference in prices to increase by 4.99 cents per gallon holding Top Tier status and

city constant. In every case, the "top_tier" coefficient is statistically significant with a modest

magnitude. This would imply that TopTier stations are perceived as having better quality gas

worth paying an additional 1.3 cents per gallon compared to non-TopTier stations holding

distance between stations and city constant. Notably, the statistical significance of the "farthest"

coefficient indicates that we cannot reject the null hypothesis of no association. We must accept

the null in this case, which is the lack of correlation between distance to the farthest station and

the difference in their prices. This supports the hypothesis that stations located far apart have

little influence on the prices they charge. Noting the significance of the main independent in each

of the other models however (<0.001), we can reasonably say that stations located near to each

other are in fact more likely to affect each other's pricing behavior.

Conclusion

This study began as a chance to investigate the reasons why gas stations can be seen

charging rates as much as 10% higher than competitors either a block away or simply across the

street. After collecting the data from way.com and reading similar studies, I pivoted to examine

the broad effect of gas station clustering. I hypothesized that more competition would lead to

lower equilibrium prices in line with the theory of supply and demand. I also hypothesized, like

others, that space and distance carry an innate cost. This cost presents itself in the ability of firms

to differentiate their prices for the same product across regions.

After conducting the first four regression analyses, I am suggesting that both the number

of gas stations nearby and the distance to the closest stations influence the price of gasoline in

the local market. The number of gas stations in a quarter and half mile radius has a sizable

negative effect on the price at the 1% level, and the distance to the closest stations, specifically

the second and third closest, has a similarly sizable positive effect on the price at the 1% level as well.

The final four models which dealt with price gaps indicate two phenomena. First, the consumer's decision after factoring in geography plays a role in firm pricing behavior. At any given time, the consumer may choose to purchase gasoline from the station they are nearest to, or they may choose to travel farther to the next closest instead assuming perfect information. If the consumer knows the prices in the market, they may find the marginal benefit of travelling to another nearby station greater than the marginal cost of the extra trip. In an effort to clear the sale, each profit-maximizing firm must consider the choices available to the consumer which are dictated by geography in this case. Regardless, it seems that very distant stations do not influence each other to any meaningful degree. I refer back to the results in the two radii regressions as well as the differences regressions when making the claim above. Each additional station in the farthest band had a weaker magnitude of effect on prices, and the "farthest" coefficient indicated a lack of correlation between distance and price differentials.

In the future, I would be interested in exploring the rural and suburban markets to complement and compare to these urban markets. Given the results of this study, which indicate that the price function includes the distances to nearby competitors and, by extraction, the prices charged by nearby competitors, I would expect similar results in the less dense markets.

Works Cited

American Petroleum Institute (2022). Notes to State Motor Fuel Excise and Other Taxes.
https://www.api.org/-/media/Files/Statistics/State-Motor-Fuel-Taxes-Charts-january-2022.pdf

Bloomberg (2024). https://www.bloomberg.com/energy

Bureau of Economic Activity (2024). Personal Consumption Expenditures by State, 2023.
https://www.bea.gov/news/2024/personal-consumption-expenditures-state-2023

Bureau of Economic Activity (2024). Regional Price Parities for States, 2023.
https://www.bea.gov/news/2024/real-personal-consumption-expenditures-state-and-real-personal-income-state-and

Capozza, D. R., & Van Order, R. (1978). A generalized model of spatial competition. *American Economic Review*, *68*(5), 896–908.
https://econpapers.repec.org/article/aeaaecrev/v_3a68_3ay_3a1978_3ai_3a5_3ap_3a896-908.htm

Eckert, A., & West, D. S. (2004). Price uniformity and competition in a retail gasoline market. *Journal of Economic Behavior & Organization*, *56*(2), 219–237.
https://doi.org/10.1016/j.jebo.2003.09.006

Pinkse, J., Slade, M. E., & Brett, C. (2002). Spatial price competition: a semiparametric approach. *Econometrica*, *70*(3), 1111–1153. https://doi.org/10.1111/1468-0262.00320

Taylor, R., & Muehlegger, E. (2025). The effects of competition in the retail gasoline industry. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5181305

Top Tier™ Gasoline Brands. https://www.toptiergas.com/gasoline-brands/

United States Census Bureau (2020). Decennial Census.
https://data.census.gov/table/DECENNIALDHCVI2020.P1?q=P1:%20TOTAL%20POPULATION

Way.com Gasoline Price Reporting. https://www.way.com/gas/stations/atlanta-ga