Testing for Spatial Differentiation: Location Choices of Gasoline Stations along Interstate Highways

Aris Avgousti*

Department of Economics
University of Wisconsin-Madison

October 4, 2010

Abstract

This paper shows evidence that location choices of gasoline retailers along interstate highways reveal incentives to spatially differentiate in order to shield themselves from competition. On interstate highways, stations cannot freely choose the distance from their competitors. They have to locate either at the same exit as other rivals, or at a predetermined distance from them. The variation in the location of interstate exits allows me to test how distances between exits can explain the level of entry at an exit. Gasoline itself is homogeneous, so theory suggests that geographic location is an important way for firms to differentiate their product and gain market power. Consistent with the theory, I find that when an interstate exit becomes more geographically isolated, the probability of entry and the number of stations at an exit increases. In addition, I find that when entering a market, stations choose to locate farther from rivals in markets with lower concentration, which supports the first finding that there are incentives for spatial differentiation. Finally, I show empirical evidence that spatial differentiation results in a higher average price and more price dispersion. The former suggests that differentiation is a mechanism for gasoline retailers to gain market power, and the latter is consistent with the theory that more differentiated firms can better price discriminate across locations.

*I am grateful to Jean-François Houde for his advice and suggestions. I also thank Jack Porter, Juan Esteban Carranza, Rebecca Lessem, Ignacio Monzón, David Rivers, Carl Sanders and Yuya Takahashi for valuable comments. *Correspondence*: cell: (357) 99436112, email: aristo@ucy.ac.cy, webpage: http://www.ssc.wisc.edu/~avgousti/

1 Introduction and Related Literature

This paper shows evidence that location choices in the retail gasoline market reveal incentives for spatial differentiation. Theory suggests there are two opposing forces that determine location choice. If a station decides to locate closer to competitors, it can "steal" customers, but at the same time it faces greater price competition. Therefore, there is also the incentive to differentiate and gain market power. I test which of the two forces is stronger using a dataset of gasoline stations' locations along interstate highways. Gasoline is fairly homogeneous with respect to non-spatial attributes, therefore physical location is a way for firms to differentiate their product and gain market power.

Testing for spatial differentiation is not trivial for two reasons. First, a firm's location choice depends on the choices of the other firms. Therefore, the degree of differentiation depends endogenously on rivals' locations. In general, there is lack of exogenous variation on the distances between firms because they can freely choose their location on a continuous space. As a result, distances between competitors and the number of competitors are jointly determined. This creates a simultaneity bias in the parameter estimates of a regression that attempts to test the effect of competition on differentiation. Second, location choice depends on demand characteristics; more firms establish at high-demand locations. The distances between firms at high-demand locations are smaller. Without controlling for demand, the researcher faces an omitted variable bias and the results will be biased in favor of clustering.

I use a novel dataset that can deal with the two problems. I look at interstate highways connecting pairs of large US cities and I observe the locations of gasoline stations at interstate exits along the highway. The nice characteristic of the dataset is that there is an initial stage in which the government decides the locations of interstate exits, and then firms decide at which interstate exit to locate. Therefore, firms choose between interstate exits that have a predetermined distance from the other exits. When the government decides the location of interstate exits, it cares about factors that the firms also care about (demand patterns), and other factors that the firms do not care about (cost of construction, safety of drivers, etc.). Conditional on traffic, local population and other demand characteristics, the distances between exits are determined exogenously to firms' choices. This allows me to test how distances between exits can explain the number of stations at an exit or the probability of entry at an exit. In addition, in my data I observe demand characteristics at the finest level. I show that without controlling for traffic counts between exits and local population around the exit the results suggest that as distance from other locations increases (in other words, as differentiation from other exits increases), the number of stations decreases. When controlling for demand characteristics, the sign flips and the results suggest that there are more stations at exits that are more distant from other exits. I find that there is a higher probability of locating at an interstate exit that is geographically more isolated.

The novel dataset allows me to test for incentives for spatial differentiation while dealing with the simultaneity bias problem and the omitted-variable bias problem. I combined the dataset from four different sources: I first connected the locations of gasoline stations with the road network (the interstate exits), and then I connected the traffic counts between exits that I collected from the different state departments of transportation. Finally, I added the demographics and the zip code business patterns from the Census. In addition, I have a completely different dataset on prices at a station level, which I am able to match with my original dataset. I use the price data to empirically test the effect of spatial differentiation on average price and price dispersion.

I find that stations prefer to establish at interstate exits that are geographically more isolated. Without controlling for local demand, I find results in favor of clustering. However, conditional on local demand, the results suggest that stations prefer to establish at exits that are farther from potential competition. In addition, I show that conditional on the location of exits and the length of the highway, stations that entered the market the last two years chose to spatially differentiate more in markets with more firms. This supports the theory that firms seek to gain market power through spatial differentiation when they face greater competition. Finally, I find that average price and price dispersion is higher on interstate highways with more dispersed locations. Stations gain market power through spatial differentiation, which allows them to set higher prices. The result on price dispersion is consistent with the theory that firms that are more differentiated can better price discriminate across locations.

1.1 Related Literature

Spatial differentiation is a special case of differentiation on the space of product characteristics. The empirical literature on product differentiation generally finds results in favor of clustering. Typically, the literature looks at how measures of competition affect measures of distance between product characteristics. A positive relationship is interpreted as incentive for product differentiation. Borenstein and Netz (1999) estimate the effect of competition (measured as market concentration) on differentiation in departure times of airlines. They find that multiple airlines on a route will locate their flights closer together in time than will a single firm controlling the same number of flights. Pinske and Slade (1998) develop a test to examine whether gasoline stations with similar contractual agreements with refiners are more likely to cluster with respect to their physical location. Because contract types are associated with certain station characteristics, they interpret their results as implying that firms with similar characteristics tend to cluster. Stavins (1995) finds evidence that a firm developing a new personal computer model will locate its model closer to existing models in the product space as the number of firms in the industry increases.

Netz and Taylor (2002) finds results in favor of spatial differentiation. They investigate the effect of competition (measured as the number of nearby firms) on spatial differentiation of gasoline retailers (measured as the average distance to the nearby firms). They find that gasoline stations spatially differentiate as a reaction to increased competition[1]. Although I show a similar result in this study, it differs from Netz and Taylor (2002) in that I am able to explain station presence at each exit based on an exogenous measure of spatial differentiation; the exogenous distances between the possible locations.

Seim (2006) and Mazzeo (2002) both investigate joint entry and endogenous product choice by explicitly modeling the endogeneity of the two. Seim (2006) looks at the market of video rental stores. She finds that their observed location choices imply that distant competitors affect payoff significantly less than nearby competitors, which suggests an incentive to spatially differentiate. However, the complicated road network inside cities makes it difficult to pin down the competitive effects in cases where demand comes mainly from moving population (vehicles) rather than nearby residing population. Houde (2009), who estimates demand for gasoline, notes: "...since consumers are not using the product (gasoline) at home, stations located anywhere along a common commuting path should require similar shopping costs."[2] Hence, nearby stations are not necessarily immediate competitors if they are not located along a common commuting path On the other hand, Mazzeo (2002) looks at quality choice of motels along interstate highways. He defines every interstate exit as a market. Like Seim (2006), Mazzeo (2002) also finds strong incentives for product differentiation.

My setting is suited to test for spatial differentiation for three reasons. First, interstate highways

[1] They define a market as a buffer zone around a station and their differentiation measure is the average distance from competitors within the buffer zone. This imposes same competitive effect from all stations within the buffer zone, regardless of distance or traffic patterns.

[2] Houde (2009) develops a model where consumers are located at a set of points in their driving path from home to work instead of at single points (e.g. home). He estimates the demand parameters and the marginal cost, and performs counterfactuals under different vertical contracts, including a ban on vertical integration.

are fairly linear markets; therefore, distances measure transportation cost of consumers as in the theoretical linear model, which in turn implies that distance is a measure of spatial differentiation. When the road network is complicated, distances do not necessarily measure spatial differentiation because potential demand for nearby firms in different roads can be very different. Second, I have information on potential demand, including local population and traffic counts between exits. Location choice depends on demand, and without controlling for it the results will be biased in favor of clustering. Third, the different locations of interstate exits are chosen by federal and state authorities, which means that firms' choice is between exits that have a predetermined distance from other exits. I expect that firms choose their location considering - among other things - the distance to other exits, because locating farther allows them to be farther from potential competition.

In the next section, I describe the location decision problem of the firms. In section 3, I describe the markets and the locations. In section 4, I estimate two equations that provide evidence of spatial differentiation. Section 5 presents empirical evidence that spatial differentiation results in a higher average price, and section 6 concludes.

2 Location Decisions

The decisions of firms about the locations of their gasoline stations along interstate highways are extremely important to the profitability of the station, and at the same time they are complex. I assume that there is a three stage-game that determines firms' choices. In the first stage, the government constructs the interstate highways and decides the locations of the interstate exits. In the second stage, gasoline stations choose to locate at interstate exits taking consumers'behavior as given, and in the last stage stations compete in prices.

2.1 Government's Decision for Interstate Exits

Although construction on the Interstate Highway System continues, the largest part of the system was constructed between 1960 and 1980. By far, most interstate exits were built during the initial construction of the interstate highways, not added at a later date, and were eligible for federal construction funds under the Federal-Aid Highway Act of 1956. At that time, the government could not perfectly forecast traffic patterns. At the same time, officials expected commercial services to be provided near exits, however, decisions on exit locations were not and are not related directly to these services. From the instruction manual of "The 1987 Estimate of the Cost of Completing the

Interstate System":[3]:

> "It is important that interchanges be so located to properly discharge and receive traffic from other interstate and Federal-aid system routes, or major arterial highways or streets. It is equally important that they not be spaced so closely as either to unnecessarily increase the cost of the System or interfere with the free flow and safety of traffic on the Interstate System."

The section continues by defining specific distances between exits[4]. In addition, the state authorities that request the construction of an additional interstate exit, are required to present - among other things - a document stating the distance to the next exit in each direction, a statement of construction cost, and environmental impacts.

It is clear that when the government decides where to locate the interstate exits, it takes into account factors that the firms also care about, like local demand and traffic patterns, and factors that the firms do not care about, like construction cost, environmental impacts, safety of drivers and geographical barriers. Conditional on demand characteristics, distances between interstate exits are exogenous to firms' location choices. Therefore, firms' location choice is between locations that have an exogenous, predetermined distance from other locations. Here, I implicitly assume that the firms' location choices do not affect demand and that there is no firm that can affect government's decision.

2.2 Demand for Gasoline

Customers are either local drivers or drivers along the highway who seek for gas when their tank is not full, or near empty. These drivers make a discrete choice of where to fill their tank. They differ in their location preferences of where to stop, their values for the outside option and the information they have about prices along the highway. Drivers have different location preferences because they have different origin-destination pairs and different amount of fuel remaining in their tank. The fuel remaining in their tank also determines their relative value for the outside option, since they may have the option to fill their tank at their final destination or within towns.

Information also plays a key role here. If no consumer is informed about prices, then only their location preferences will determine where they stop. In this case, the demand at each interstate

[3]From the section "Interchange Spacing".

[4] Exits should not be closer than an average of 2 miles within urban areas, not closer than an average of 4 miles in suburban areas and not closer than an average of 8 miles in rural areas. In urban areas, the minimum distance should not be less than 1 mile, and in rural areas not less than 3 miles.

exit is given exogenously to firms' price decisions. Therefore, each interstate exit is the relevant market. The incentives to differentiate along the highway segment by being a monopolist in the local market are higher. On the other hand, if consumers are informed about prices they choose their stop according to the utility they receive from location preferences and price. Information creates price competition between stations located at different exits. In this case the whole segment of the highway connecting two cities is the relevant market. The incentives to differentiate along the highway are lower than in the case of no information, but they can still be important when location preferences of consumers are strong such that they provide enough market power to differentiated stations.

For example, most tourists have no information on prices along a highway, but local drivers and professional (truck) drivers do. When long-distance drivers are uninformed and local commuters are unlikely to buy along the highway, then each exit is a local market and the incentives to differentiate along the highway are very strong. If long-distance travelers are regular drivers along a route or professional drivers, then price competition between stations at different exits is likely to be strong. The incentives for firms to spatially differentiate are lower than in the case of no information.

2.3 Stations' Location Decisions

Gasoline stations take the position of interstate exits and the distribution of drivers along the highway as given, and then decide at which exit to locate. Relocation is prohibitively costly. The stations care about local customers, the number of long-distance travelers, local entry costs, and the degree of competition at each location. Local customers have a better outside option and they are better informed about prices than long-distance travelers. Stations compete with other stations in the same location, and, in cases in which demand from local commuters is significant, they compete with stations across different locations along a highway too. Therefore, when stations choose their interstate exit they often care about the distance from other exits. Conditional on demand characteristics, these distances are predetermined and exogenously chosen by the government, in the first stage. Therefore, I can test how distances between exits can affect the number of stations at an exit.

In areas with large population there are more interstate exits and as a result exits are closer to each other. At the same time, more stations are located at these exits because of the favorable local demand conditions. This implies that without controlling for local demand the results will suggest that stations are located at exits that are closer to each other.

In addition, it is important to control for long-distance travelers too. These customers have a lower value for the outside option and are not as well informed about prices as commuters. Therefore, when there are more long-distance travelers relatively to local commuters, the incentive to spatially differentiate is higher because the exits are closer to local markets and they provide more market power than in the case in which long-distance travelers are relatively fewer. Controlling for the number of long-distance travelers is important in order to identify the different incentives to spatially differentiate. Here, I assume that long-distance travelers do not solve an optimal stopping problem; they do not observe all the prices along the highway and it is too costly to get this information. Rather, they follow a heuristic rule of stopping at the position where their tank is near empty. Therefore, firms' location decisions cannot affect the distribution of "positions" of long-distance travelers along the highway (because they cannot affect the position where the drivers have their tank near empty). What firms are able to do is to exploit the incomplete information of long-distance travelers by spatially differentiating. With spatial differentiation, stations gain market power.

2.4 Price Competition

Gasoline retailers post their prices and they can easily adjust them. Therefore, they can easily response to a change in price from an adjacent competitor. Consumers can compare prices of adjacent stations without leaving their car. Therefore, price competition is particularly strong for gasoline stations at the same interstate exit and subsequently, the incentives to spatially differentiate are strong too.

After firms choose locations, they take into account consumers' preferences and rivals' locations and engage in price competition that results in a Nash equilibrium. For most stations, entry and location decisions are made many years ago and are costly to change. The number of stations and their locations are fixed when price decisions are made, and firms are not able to perfectly predict prices in 5, 10 or 20 years ahead. This implies that the equilibrium price vector should depend on the market geography, but the location decisions cannot change based on today prices.

3 Data

The retail gasoline market is an industry with a homogeneous product and many retailers with small market shares. In 2008, the sales of regular motor gasoline to end users was about 65.77 billion dollars.[5] This paper focuses on gasoline retailers along interstate highways. Interstate highways offer a nice example to study spatial competition since it is a relatively simple road network without complex traffic patterns and a known line of travel between two points. A market is defined as a segment of an isolated interstate highway connecting two large cities. A market contains a set of interstate exits and stations located at the exits. Stations compete for drivers traveling on the highways and for local customers.

3.1 Markets and Locations

The Interstate Highway System is a network of limited-access highways in the United States, with a total length of $46,876$ miles. It was authorized by the Federal-Aid Highway Act of 1956, and most of the existing highways and exits were constructed between 1960 and 1980. This freeway system serves nearly all major U.S. cities, with many interstates passing through downtown areas. The distribution of virtually all goods and services involves interstate highways at some point[6]. The vast majority of long-distance travel, whether for vacation or business, uses the national road network[7]. Of these trips, about one-third (by the total number of miles driven in the country in 2003) use the Interstate System[8]. Commercial services for travelers have flourished along interstate highways, even in areas with little local demand.

Markets are selected such that long-distance travelers can substitute between the gasoline stations, and so that they are isolated from competition outside the market definition. The initial set of cities is the set of Metropolitan Statistical Areas (MSAs) and cities with population higher than $120,000$. Markets are defined as the shortest paths between each pair of cities[9]. Paths are selected to satisfy the following criteria:

1. They are on the Interstate Highway System.

[5] It is about 0.44 percent of the USvgross domestic product (GDP).

[6] Caltrans (2006). "The Interstate Highway System Turns 50".

[7] "Table 1-36: Long-Distance Travel in the United States by Selected Trip Characteristics". Bureau of Transportation Statistics. 1995.

[8] "Annual Vehicle Distance Traveled in Miles and Related Data". Federal Highway Administration. 2003.

[9] Data on the US road network is obtained from the ESRI StreetMap dataset, which is an extension tool of the mapping software ArcGIS. It provides a nationwide street network that can be used for map visualization, geocoding of addresses and shortest path calculations.

2. They are isolated from other large cities and other limited-access highways.

3. The driving time between cities ranges from 60 to 300 minutes.

Finally, I drop tolled highways and markets with no traffic information[10]. The criteria aim to produce homogeneous markets. The first criterion produces fairly linear markets with a discrete number of locations and information on traffic counts. I impose criterion two in order to avoid outside competition from gasoline stations in other cities or along other limited-access highways. The third criterion makes sure that the number of long-distance travelers is not small and that they do not need to fuel more than once. These refinements resulted in 238 markets. Figure 1 presents the mapping of the selected markets on the US landscape with the dots representing the locations of gasoline stations along the interstate highways. The selected markets cover almost entirely the continental United States.

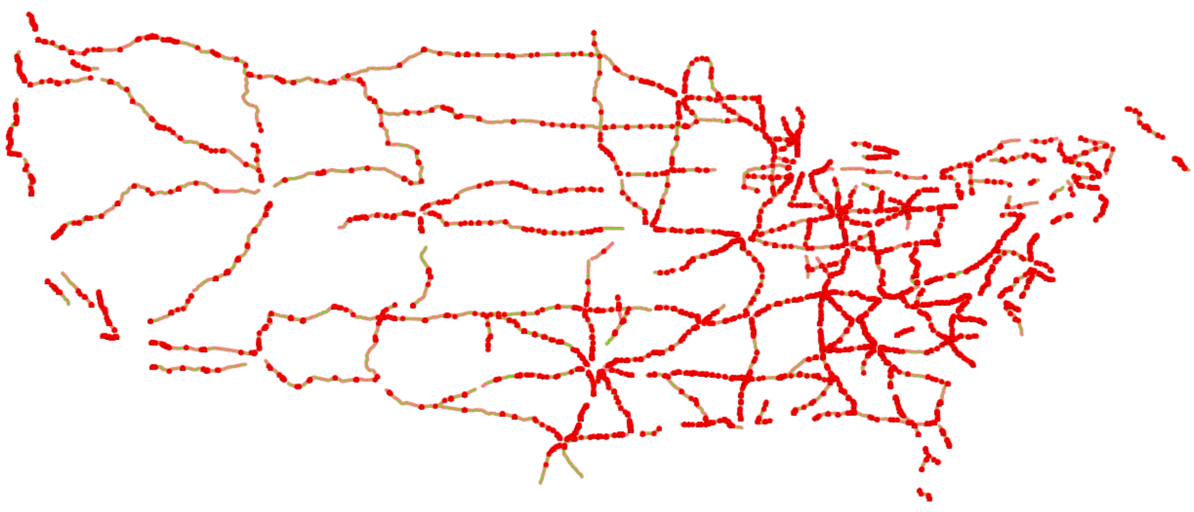

Figure 1: Selected Markets

Locations are defined as a buffer zone of half a mile around the interstate exits, but results are robust to the definition of a one mile buffer zone around the exit. The data on gasoline stations is obtained from the December 2007 version of *ReferenceUSA*. This is an online business directory that contains information for more than 14 million businesses in the US. Data on geographic coordinates, address, chain affiliation, lines of business and year entered the dataset is available. For each exit, I observe traffic counts, local population, other local characteristics, and relevant demographics of the exits' census areas[11]. Traffic counts are average annual daily traffic (AADT) in 2007,

[10] I drop 19 markets with no traffic information, only within specific states.

[11] I use local population of block groups within three different ring zones of the exits: 2 miles, 2 to 5 miles and 5 to 8 miles. The census areas of the exit are block group, census tract, zip code and county.

and information is obtained from the individual state departments of transportation. They measure the total number of vehicles that drive by an exit for both directions of flow, but not necessarily exited at that exit. Demographics are obtained from the Census Bureau.

Table 1 presents summary statistics on interstate exits. A striking characteristic is that around 57 percent of the exits are not occupied. I proxy demand by traffic counts and local population. On average there are $31,000$ vehicles per day passing through an exit. The average population of block groups within two miles of an exit is $8,700$[12]. There is considerable variation in both of these characteristics. In addition, I construct a dummy variable indicating whether the exit is an outlet to another important highway[13]. I use median contract rent as a proxy for business rent, which is a fixed cost for gasoline stations. In addition, I use the percentage of occupied houses that are rented at the block group level to proxy for entry cost and zoning restrictions. As the percent of rented houses increases, I expect that entry cost will decrease and that zoning restrictions will be weaker[14]. Entry decisions may also be affected by state gasoline taxes, since they do not change often and the ranking of low to high tax states is relatively constant over time. Finally, I include information on nearby hotels and restaurants. Long-distance travelers often combine eating, fueling and resting, so demand at exits with more restaurants and more hotels may be higher.

Table 2 presents descriptive statistics on markets. There is significant variation in the number of interstate exits and gasoline stations across markets. I distinguish firms from stations by defining a firm as a collection of stations that have the same refiner affiliation. Therefore, the number of firms is the number of brands plus independently owned stations. The number of firms operating a fixed number of stations is a measure of market concentration. I expect greater price competition in a market with ten firms operating ten stations than in a market with two firms operating ten stations. I do not observe the number of long-distance travelers, but to proxy for this variable I use the minimum of the traffic counts over locations. This can be thought as an upper bound for long-distance travelers[15].

[12] A census block group is a geographical unit used by the US Census Bureau and its size is between the census tract and the census block. It is the smallest geographical unit for which the bureau publishes sample data. The United States has $211,267$ block groups.

[13] Important highway is a class 2 road as defined from the Federal Highway Administration. This definition includes most of the state highways and some county highways.

[14] This is also used in Netz and Taylor (2002). Being easier to rent in the area indicates that you can more easily establish in the area. In addition, I expect that in a fully owner-occupied area residents are more strongly opposed to allowing business establishments, or they have more lobby power to put zoning restrictions.

[15] I do not observe the number of vehicles that exit and enter the highway from each exit. I observe, for example, that between Exit 6 and Exit 7 there are $10,000$ vehicles, and between Exit 7 and Exit 8 there are $9,000$ vehicles. I know that $1,000$ more vehicles have exited at Exit 7 (but I do not know if $1,000$ vehicles exited and 0 entered the highway, or if $5,000$ vehicles exited and $4,000$ entered, or something different). Hence, I use $9,000$ to proxy for long-distance travelers,

Figure 2 presents an example of a market. Des Moines and Iowa City are connected by Interstate 80. The market is the whole segment of the highway. All the interstate exits on the segment are possible locations (large dots on the line). A station is defined as a competitor in the market if it falls within half a mile from the exit ramps. Under this definition, there are 10 stations in the market (marked by the underlined numbers that indicate their exit number) and 20 possible locations.

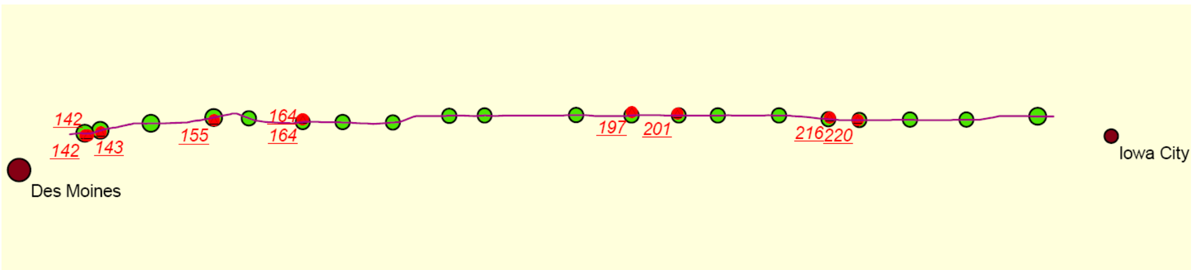

Figure 2: Interstate 80, connecting Des Moines, IA and Iowa City, IA

Table 3 reports summary statistics for the gasoline retailers in my data. More than 60 percent have a convenience store. The percentage of independently owned stations is around 40 percent, and from the remaining brand affiliated stations almost half of them are affiliated with the major, international brands (Shell, BP, ExxonMobil, Chevron, Texaco-Conoco). An interesting statistic is that a gasoline station faces almost 2 competitors on average at the same interstate exit, even if the average number of competitors at an exit is less than one. Figure 6 shows the distribution of the number of stations over exits.

3.2 Location Characteristics

A first look at the data suggests that location choices are mainly driven by traffic patterns. This section presents two examples of markets with different traffic patterns and different location choices. Figure 3 shows the average daily traffic and the number of stations at all exits between Alexandria, LA and Lafayette, LA, on Interstate 49.[16] The exits are sorted on the horizontal axis using their normalized location on a line of length one, from Alexandria to Lafayette. The stations are located at exits where traffic is higher. Almost all stations are at exits where daily traffic counts are more than $30,000$. Even though many other factors affect location choice, such as local population, entry cost

but this is actually an upper bound. Other proxies, like the 10th percentile of traffic over locations, or the 25th percentile, provide similar results.

[16] The market has 20 exits-locations, 90 miles distance between the two endpoint cities and 88 minutes driving distance.

and zoning restrictions, traffic counts seem to drive location choices in this market.

Figure 4 shows the traffic and the number of stations at the sorted exits between Bismarck, ND and Fargo, ND, on Interstate 94.[17] Traffic follows a similar distribution as the previous example, but the magnitude is less. Traffic ranges from around 6,000 to 17,000 counts. In this market, most stations establish roughly in equal distance from one another. There are only two cases in which stations establish close to each other, and both of these cases can be explained by higher local population (Figure 5).

The gasoline stations in Figure 4 are more equally spaced than the gasoline stations in Figure 3. In other words, spatial differentiation along Interstate 49 between Alexandria and Lafayette is lower than differentiation along Interstate 94 between Bismarck and Fargo. In cases in which local demand is high (Figure 3), the demand effect is stronger than the market power effect and stations cluster. In cases in which demand is low (Figure 4), we can observe a preference of stations for differentiation by being more equally spaced from one another. The variation of exit locations and of the number of stations at each location, allows me to examine the effect of exit geographic isolation on the probability of entry and the number of stations, while controlling for traffic and local population.

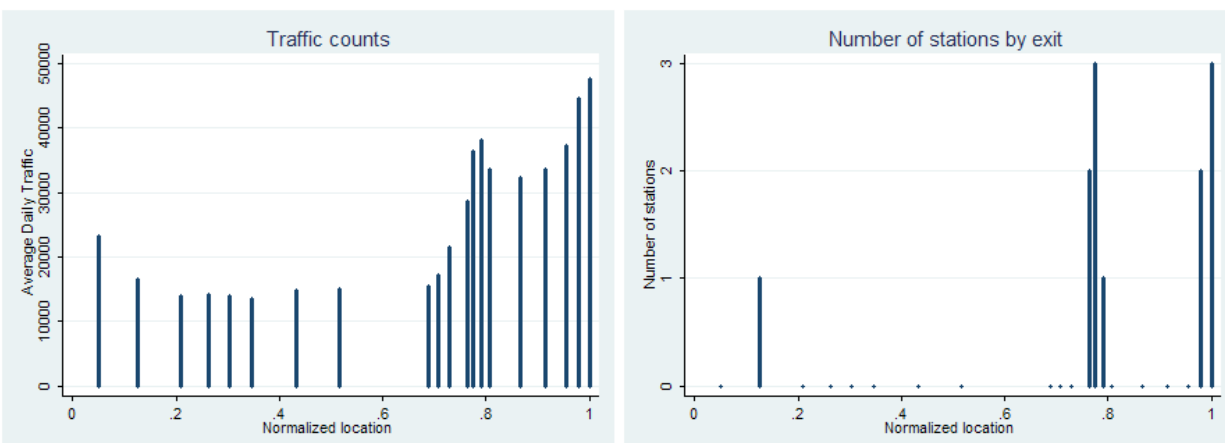

Figure 3: Alexandria, LA - Lafayette, LA (Interstate 49)

[17]This market has 50 exits-locations, 181 miles distance between the two endpoint cities and 183 minutes driving distance.

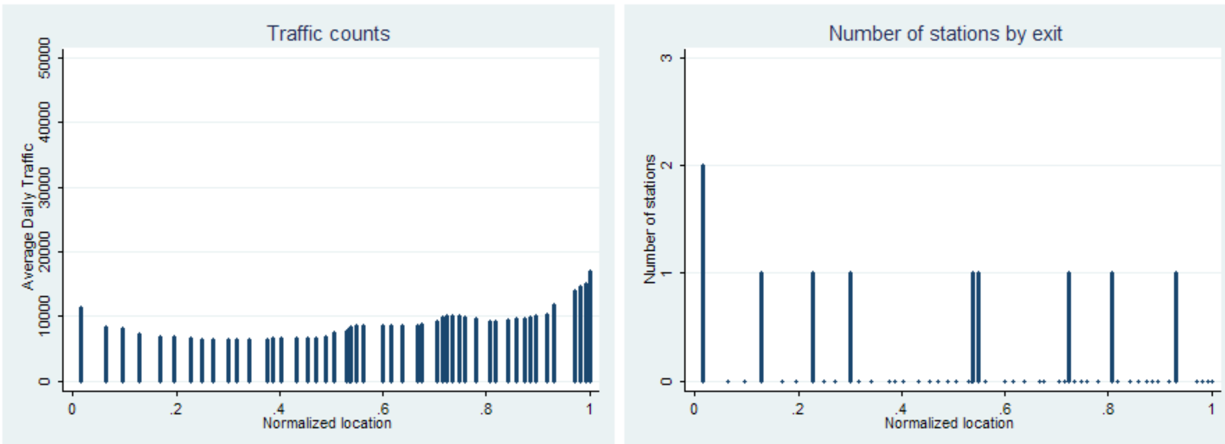

Figure 4: Bismarck, ND - Fargo, ND (Interstate 94)

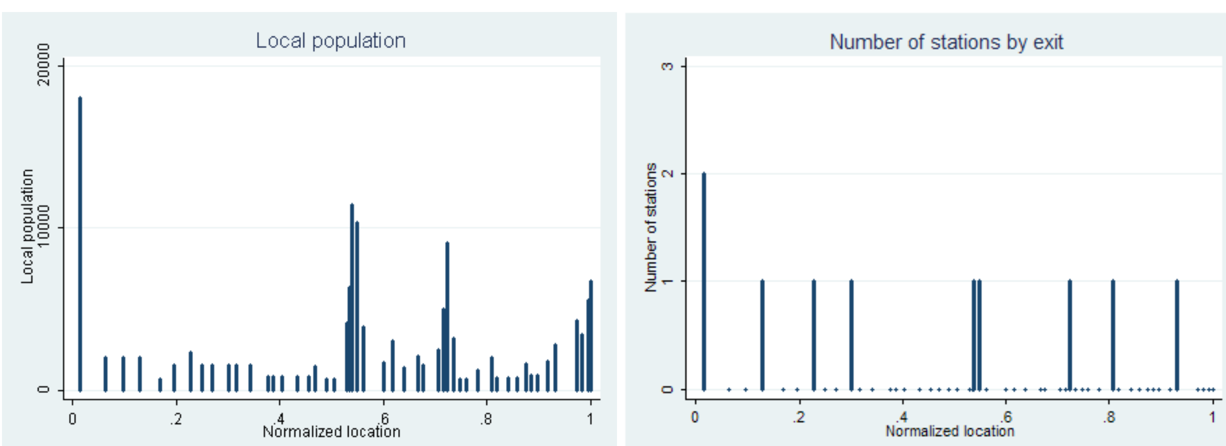

Figure 5: Bismarck, ND - Fargo, ND (Interstate 94)

4 Testing for Spatial Differentiation

This section provides evidence that gasoline stations on interstate highways have incentives to spatially differentiate. I estimate two different equations. First, I am interested in whether stations prefer to establish at interstate exits that are farther from other exits, and therefore farther from potential competition. Then, I estimate a different equation of spatial differentiation of the stations that entered the market the last two years. I am interested in whether entrants differentiate more in markets with a lower degree of market concentration, and hence, a higher degree of expected price competition.

4.1 Measures of Spatial Differentiation

Even though interstate highways are relatively linear markets, it is still indeterminate which measure captures differentiation from other exits in the market. Finding the appropriate measures of differentiation has not been an easy task in this literature and it often depends on the characteristics of the markets under consideration. In particular, we know that a natural measure in the case of geographic differentiation is geographic distance. However, which distance? Average distance from competitors? weighted distance (if yes what is the weight)? total distance? distance to the closest competitor? or something else? All of the above measures, perhaps with the exception of total distance, have been previously used in the literature.

I use differentiation measures that are more appropriate for linear markets and are either implied by the theory on spatial differentiation, or used in the previous literature, or suggested in this chapter as I believe they are better suited for the specific application I consider. First, I describe a measure that accounts for the closest location from each side. Then I describe a measure of geographic proximity to all other locations, and finally I describe a measure previously used in the literature. The results presented in the next section are robust to the different measures I use.

On a continuous line, spatial differentiation is measured as the distance to the closest location. Differentiation along interstate highways can be parallelized with differentiation along a line with discrete locations. Therefore, an appropriate measure of spatial differentiation is the average distance to the two closest locations from each side:

$$Isolation2_j = \frac{1}{2}\left(|d_{j,left} - d_j| + |d_{j,right} - d_j|\right) \tag{1}$$

where $d_{j,left}$ and $d_{j,right}$ is the position of the closest exit to the "left" side and "right" side respectively.

I check the robustness of the results to the selection of this measure, by using other measures of differentiation. If the number of different locations are relevant, then we can use a measure that accounts for the number of possible exits and also for how far away these exits are. I call this a measure of geographic proximity to other exits in the market. I define proximity for exit j as the weighted sum of other exits in the market, weighted by the reciprocal of the Euclidean distance between exit j (at position d_j, which is defined as the distance from the first exit after the first endpoint city) and each other exit (at d_k):

$$Proximity_j = \sum_{k \neq j} \frac{1}{|d_k - d_j|} \tag{2}$$

This measure puts less weight to exits that are farther away. It is affected by the number of exits on the highway but also the distance between exits. It is important to account for the number of possible locations that potential competition can establish, therefore I use a sum over all exits in the market. At the same time, an exit that is closer should have a bigger effect than an exit farther away[18]. A third measure I use is the proximity measure in the above equation (equation (2)) weighted by the reciprocal of the *squared* Euclidean distance.[19]

If all other exits are important, another appropriate measure is the average weighted distance to all the other exits:

$$IsolationALL_j = \frac{1}{(J-1)} \sum_k \sum_{k \neq j} \left[|d_k - d_j|\right]^{\alpha},\ 0 < \alpha < 1. \tag{3}$$

This measure is used by Borenstein and Netz (1999), who estimate the effect of competition (measured as market concentration) on differentiation in departure times of airlines. Their differentiation measure is a measure around the clock; it is the weighted average distance from other flights, weighted such that the significance of flights that are farther away is lower. They estimate the model using different values of the significance parameter, α, and I follow the same robustness test here.

[18]For example, an exit one mile farther adds 1 unit to proximity, while an exit 100 miles farther adds 0.01.

[19]Third measure: $Proximity2_j = \sum_{k \neq j} \frac{1}{(d_k - d_j)^2}$.

4.2 Entry of Stations and Geographic Isolation of Exits

I test whether firms are attracted to locations that are geographically more isolated from other locations. The number of stations at an exit depends on local exit characteristics and the distance to the other exits. If stations prefer to locate at exits that are farther from other exits, then I should find a positive relationship between the number of stations and the distance measure of geographic isolation.

I estimate the following equation using least squares:

$$N_{jm} = X'_{jm}\beta + \gamma Isolation_{jm} + \varepsilon_{jm} \tag{4}$$

The dependent variable is the number of stations at the exit level or an indicator if the exit is occupied. The vector X_j includes local demand variables and cost characteristics around exit j. I expect that stations prefer to locate at areas with higher demand and lower entry cost. $Isolation_j$ is a measure of geographic differentiation from other exits and is defined as in equation (1). If more stations prefer to locate at isolated exits, I should find that γ is positive.

The parameter γ should only capture the effect of spatial differentiation and should be free of any demand effects. In particular, exits that are geographically more isolated may attract a higher number of long-distance travelers, simply because travelers have no other nearby options. In other words, if the location preferences of highway drivers follow the uniform distribution, then geographically isolated exits will be the closest exits for more drivers than geographically clustered exits. This implies that geographic isolation, $Isolation_j$, will also include a demand effect. To account for this, I develop a simple model in which I use the traffic counts along the highway to predict the number long-distance travelers that stop for gas at each exit.

The distribution of stops of long-distance travelers along the highway is assumed to follow the uniform distribution. Therefore, exits that are geographically more isolated will have a higher number of predicted stops of long-distance travelers.[20] The initial estimate of long-distance travelers is the minimum traffic over exits, and then the assumption about the uniform distribution is used to predict the number of long-distance that stop at each exit. Therefore, I construct a new variable for the predicted number of drivers exited for gas at each exit, that accounts for the possibility that geographically more isolated exits attract more long-distance travelers.

[20] The uniform distribution is assumed to be between 0 and 360, which comes from a maximum driving distance of 360 miles with a full tank. I use an average tank capacity of 15 gallons and an average gas mileage of 24 miles per gallon.

Before estimating equation (4), I need to address some econometric concerns. Unobserved market characteristics can be correlated with the number of stations and can also be correlated with right hand side variables. Hence, the residuals will depend on a market-specific component, μ, and an idiosyncratic component, η. The residuals can be written as:

$$\varepsilon_{jm} = \mu_m + \eta_j \tag{5}$$

I include market dummies to account for any unobserved market differences. Furthermore, if X also has a market-specific component, the residuals within a market will be correlated even when using market dummies. The OLS standard errors will understate the true standard errors. To correct for the correlation of the residuals, I compute clustered standard errors within markets[21].

Table 4 shows the estimated coefficient of $Isolation$, γ, as I account for more demand characteristics. The left column indicates the control variables included in each specification of equation (4), and the right column shows the estimated γ as the specification changes. The first row shows the coefficient of $Isolation$ without controlling for demand characteristics. In this case, the coefficient is negative, suggesting that less stations will locate at geographically isolated exits. However this result changes as I include demand characteristics. When I control for traffic and predicted potential demand, in the second row, the sign of $Isolation$ remains negative but becomes statistically insignificant. In the third row, I control for local population and the sign becomes positive and statistically significant. Local population is the factor that changes the sign of γ. In the last two rows, I add controls for entry cost, number of hotels,number of restaurants, distance to nearest city and demographics of the nearest city. The magnitude of the isolation parameter becomes even stronger. Without controlling for demand characteristics, I am not able to identify the preference for establishing at isolated exits. At areas with large local population, the number of exits is higher and as a result exits are closer to each other. This implies that without controlling for local demand, I find that stations are located at exits that are closer to each other. When testing for product differentiation, researchers should control for the size of demand to avoid biased estimates.

Table 5 shows the estimated coefficient of $Isolation$ using a different sample size or a different estimation method. Column 1 shows the same least squares estimates of γ from Table 4 for comparison purposes, and column 2 shows the least squares estimates of γ when the sample is limited only to occupied exits. If there are unobserved zoning restrictions or unobserved costs at the non-

[21] The squared sum of $X_{jm}\epsilon_{jm}$ is assumed to have the same distribution across the markets. Thus clustered standard errors are consistent as the number of markets grows (Donald and Lang (2007) and Wooldridge (2003)).

occupied exits in urban areas, then these exits will have zero stations and high proximity to other exits, driving the positive coefficient of $Isolation$. The estimated coefficient in column 2 follows the same pattern as the one in column 1, suggesting that the concern stated above is rejected from the data. In column 3, I show the probit estimate of γ when the dependent variable is an indicator if the exit is occupied. If there are cases where an isolated exit has many stations located there only because they did not have another nearby location choice, it will result in a positive γ that cannot be interpreted as evidence of a preference for spatial differentiation. The probit estimates show that this is not driving the results; γ is positive when controlling for local demand.

Table 6 presents the least square parameter estimates for different measures of geographic isolation/proximity. Column 1 shows the estimate for isolation as in equation (1). In column 2, I use the measure of geographic proximity as defined in equation (2). In column 3 proximity is as in equation (2), but weighted by the reciprocal of the *squared* Euclidean distance. It is clear that the sign of the coefficients of all these measures follow the same pattern. After controlling for traffic and local population, the results suggest that stations establish at geographically isolated locations.

Table 7 shows the least square estimates for selected parameters, using $Isolation2$. In column 1, the location is defined as a single exit. In column 2, the location is a collection of exits within one mile of each other. The second column is a robustness check, to show that the very nearby exits with fewer stations than isolated exits do not drive the result.[22] The results are robust to the different definitions of location. Next, I focus on the results of column 1.

The number of stations at an exit is negatively affected by the geographic isolation from other exits. $Isolation2$ is an index and the magnitude of the coefficient depends on the scale of the variable, making it difficult to interpret its magnitude. The coefficients for the demand variables have the expected sign. They suggest that stations tend to cluster when the volume of local traffic is high, or when the predicted stops of long-distance travelers is high, or when the exit is an outlet to an important highway. Local population also has a strong positive effect to the number of stations. The presence of hotels attracts more stations, suggesting that consumers like to combine their services. The entry cost variables also have the expected sign. I assume that a higher ratio of rented houses indicates a lower entry cost and fewer zoning restrictions. The estimates suggest that firms are deterred from locating at areas with higher entry cost, more zoning restrictions and

[22] Assume two different cases: 1) a very isolated area with two very close by exits (within one mile) that have one station each, 2) a similarly isolated area with one exit that has two stations. The second exit will have more stations and higher degree of isolation, therefore it will seem that stations choose to be differentiated. In the first case, even though stations are similarly differentiated, it will seem that there are less differentiated. The results of column 2, suggest that this is not driving the results.

higher gasoline taxes.

4.3 Other Interpretations

The above results show that as an exit becomes more geographically isolated, it attracts more stations. I interpret this as revealing incentives for firms to spatially differentiate along the highway. However, a higher number of stations at a geographically isolated exit can also be considered as a result of stations' preference to cluster around this exit. This interpretation is the opposite from the one presented in this paper. A cluster of stations within an exit may result in strong price competition that will offset any benefits from being more spatially differentiated from other locations along the highway.

I address the above concerns, by reporting additional results that support the first interpretation of incentives for spatial differentiation. First, the results of the probit model in Table 5 suggest that as an exit becomes more geographically isolated, the probability of being occupied by at least one station increases. About 46 percent of the occupied exits are one-station exits and 73 percent have two stations or less, which implies that there are only a few cases in which we observe clustering of a large number of stations at an exit (only 5 percent of occupied exits have more than 4 stations, or only 2 percent of all exits). Therefore, the indicator for entry at an exit mainly comes from the exits with two or fewer stations. The result is not driven by a large number of stations at various exits.

Next, I show that the observed geographic dispersion of stations, is higher in more competitive markets. In other words, we observe stations to be more equally spaced along the interstate segment when the market is more competitive. This minimizes the concerns that the observed stations locations are a result of a preference for clustering.

4.4 Spatial Differentiation of New Entrants

This subsection provides supporting evidence of spatial differentiation of gasoline stations on interstate highways. I follow a similar procedure to the literature of spatial differentiation by examining how new entrants choose to spatially differentiate when facing a different number of competitors. I assume that a new wave of entrants choose locations, taking location choices of incumbents as given. Then, they have to compete in prices for long-distance travelers and local consumers. Here, I distinguish the number of firms from the number of stations. While the number of stations is the number of retailer establishments, the number of firms is the number of brands plus the number of independents. I assume that there is a higher degree of price competition when there are more firms operating the same number of stations. Therefore, competition is measured as the number of firms in the market, controlling for the number of stations. If the market power effect is stronger than the effect of "stealing" customers, entrants will locate farther away from other stations as a response to increased competition, and I should find a positive effect of the number of firms on spatial differentiation. On the contrary, if the effect of "stealing" customers is stronger, entrants will locate closer to existing stations when competition increases, and I should find a negative effect.

Spatial differentiation for entrant i is calculated as in Borenstein and Netz (1999):

$$Diff_i = \frac{1}{(n-1)} \sum_{k \neq i} \left[|d_k - d_i| \right]^{\alpha}, 0 < \alpha < 1 \tag{6}$$

When entrants choose a location with many stations or with many stations at nearby locations, this measure is small. On the other hand, when most of the entrants are located at exits that are farther from incumbents, this measure is large. The parameter α represents the significance of stations that are farther away. If α is close to 1, then the index is equally affected by smaller and larger distances, and is very close to the simple average. If α is close to 0, then the index is more strongly influenced by the closer stations. I estimate the model using $\alpha = .5$, however, the results are similar for $\alpha = .2$ and $\alpha = .8$.

Average differentiation in the market, $AvgDiff$, is the average of spatial differentiation of each entrant:

$$AvgDiff = \frac{1}{n(n-1)} \sum_{i} \sum_{k \neq i} \left[|d_k - d_i| \right]^{\alpha}, 0 < \alpha < 1 \tag{7}$$

Borenstein and Netz look at differentiation of airlines in the departure times of their flights. The index in (6) measures differentiation around the clock. In the case of a linear market, even if the stations are equally spaced, the measure is higher for stations located on the sides of the market. Therefore, it is possible that in markets with a higher fraction of exits near the extremes there are more stations located near the extremes, in which case average differentiation will be higher. I control for this with an index calculated as in (7), for exit isolation from other exits in the market (instead of station isolation from other stations).

Borenstein and Netz normalize the average distance by the maximum possible differentiation, so that they can compare differentiation across samples with different number of flights. Maximum possible differentiation is calculated as $AvgDiff$ in the hypothetical case in which flights are maximally dispersed. Similarly, I assign stations such that they are maximally dispersed in the theoretical sense (equally spaced from each other), and then calculate $AvgDiff$. My setting is not on a continuous space where firms can freely locate. To get a measure of maximum dispersion, I first have to assign the stations at exits so that they are more dispersed. This problem is called the discrete p-dispersion problem. It is not simple, since there are different networks of exits in different markets. In addition, the number of stations is often significantly different than the number of locations, so calculating the measure for all possible combination of location choices is impossible[23]. Instead, I use an approximate method to assign stations such that they are close to maximally dispersed. I outline the algorithm I use in Appendix 6.[24]

Then, I calculate spatial differentiation for each station as in equation (6), for the maximally dispersed stations. The measure of theoretical maximum differentiation is calculated as the average distance of the maximally dispersed stations using equation (7). Let this measure be $DispDiff$. Then, I define market spatial differentiation, $MarketDiff$:

$$MarketDiff = \frac{AvgDiff}{DispDiff} \tag{8}$$

$MarketDiff$ is a normalized measure of average differentiation of new entrants. The measure is normalized by the theoretical maximum dispersion, so that it is not affected by the number of gasoline stations along a highway. In other words, this measure gives the ratio of how different is observed differentiation to differentiation of maximally dispersed stations (evenly spaced stations). This allows for comparisons of differentiation across markets with a different number of stations.

I am interested in whether entrants are more dispersed in markets with a higher number of competitors. I estimate the following equation, using least squares:

$$MarketDiff_m = X'_m\beta + \gamma N_m + \epsilon_m \tag{9}$$

The dependent variable is differentiation of stations that entered market m the last two years,

[23] For example in a market with 60 locations and 20 stations, I have to calculate the differentiation index for $4.1918445 \times (10)^{15}$ different configurations, and then select the maximum.

[24] The main idea is that I define "fake" locations of stations such that they are maximally dispersed on a continuous line, and then sequentially assign the stations to the exit that is closest to their "fake" position, as long as the exit is not occupied. Then, I sequentially assign the remaining stations, not assigned in the first step, so that each time I fill the non-occupied exit that is between the longer gap of occupied exits.

and is defined as in equation (8). In the data, I do not observe the year that each station was established, but I observe the year it entered the dataset, which can be either the year of establishment or the year it changed ownership. Therefore, I use the year the stations entered the dataset to calculate spatial differentiation for the stations that entered the last two years (2006 and 2007). The vector X includes market characteristics, such as the number of long-distance travelers, average traffic, average local population and the length of the market. N is the number of firms in the market, which is defined as the number of brands plus the number of independently owned stations. For example, a market with 10 stations all controlled by ExxonMobil will have one firm (and 10 stations). A market with 10 stations each owned by a different firm will have 10 firms (and 10 stations). In addition, I control for exit geographic isolation using the same differentiation measure as in equation (7), but for exits rather than stations. If stations are more dispersed in response to stronger competition, I expect to find a positive effect of the number of firms on the degree of spatial differentiation. On the contrary, if stations cluster as competition gets stronger, I expect to find a negative effect.

Table 8 reports the results of estimating equation (9). The marginal effect represents the average percentage change in differentiation in response to a 10 percent increase in the independent variable. I find a positive correlation between the number of firms and the degree of spatial differentiation, implying that entrants differentiate more when competition is stronger. If the number of firms operating a fixed number of stations doubles, then the degree of differentiation of entrants increases by 15.6 percent.

In addition, the results show that long-distance travelers induce gasoline stations to differentiate. The estimated coefficients of the first two variables in Table 8, suggest that for exits with traffic around the sample average, as the number of long-distance travelers increases by a thousand, differentiation of new entrants increases by 0.2 percent. So, observed differentiation will be a little closer to the differentiation of maximally dispersed stations. If long-distance travelers increase by ten thousands, differentiation of new entrants increases by 3.2 percent. Long-distance travelers is a proxy for tourists or differently for consumers that are uninformed about prices. These travelers have a higher cost of stopping and searching for cheap gas. When they become relatively more important to firms than local demand, firms have an incentive to locate at different locations in order to avoid price competition and take advantage of the lack of information and the lower price elasticity of these travelers.

On interstate highways, firms cannot freely choose the distance from their competitors. They have to locate either at the same exit as other rivals, or at a specific distance away. The exogenous variation of the distances between interstate exits, allows me to study their effect on firms' location choices. Gasoline stations are attracted to places with higher demand, but when I control for local demand, the results suggest that stations establish at locations that are farther from other possible locations and therefore farther from potential competition. Firms differentiate in space in order to soften price competition and gain market power.

5 Spatial Differentiation and Prices

I next study how spatial differentiation affects average price and price dispersion. If product differentiation is a mechanism used by firms to soften price competition, I should observe higher gasoline prices in markets with higher differentiation. The effect of product differentiation on price dispersion is uncertain and depends on other market characteristics, like search frictions.

For most stations, entry and location decisions are made many years ago and are costly to change. In the retail gasoline market exit cost is high and relocation cost is prohibitively high[25]. Given that the number of stations and their locations are fixed when price decisions are made, and assuming that firms cannot perfectly predict prices in 5, 10 or 20 years ahead, simultaneity of prices and location choices is not a problem.

Data on prices comes from a private vendor. There is information on gasoline and diesel prices for ten days in April, 2009. Prices were collected either from reports from most of the leading chains or from data from credit card transactions. I am able to match about 90 percent of the stations in my original dataset with gasoline prices and about 57 percent with diesel prices. The unmatched stations do not have any systematic differences from the matched stations in terms of geographic location, chain affiliation or other lines of business offered. I use one gasoline and one diesel price for each station which is calculated as the average price over the ten days. Table 9 reports the market average price per gallon of gasoline and diesel and the coefficient of variation (standard deviation divided by the average price).

5.1 Spatial Differentiation and Average Price

Markets in which stations are more dispersed, should offer the ability to these stations to set higher prices. This section is concerned with the effect of spatial differentiation on average price. Theory

[25] Stations are legally responsible for the environmental cleanup of the site.

suggests that we should observe a higher average price in markets in with more differentiation. To test this, I estimate the following equation:

$$\text{AvgPrice}_m = X_m'\beta + \gamma N_m + \delta MSpatDiff_m + \epsilon_m \tag{10}$$

AvgPrice refers to market average price. The vector X includes market characteristics and N is the number of firms, which equals the number of brand affiliations plus the independently owned stations. Note, that the number of stations is different from the number of firms; the number of stations is simply the number of retailer establishments. I assume that there is lower market concentration and therefore a higher degree of price competition when there are more firms operating the same number of stations. Therefore, the number of firms controlling for the number of stations is a measure of market concentration. $MarketDiff$ is spatial differentiation for *all*[26] stations in the market, as defined in equation (8). That is, $MarketDiff = \frac{AvgDiff}{DispDiff}$, where $AvgDiff = \frac{1}{n(n-1)} \sum_i \sum_{k \neq i} [|d_k - d_i|]^{\alpha}$ (equation (7)) for the parameter of significance of farther away stations α.

Table 10 presents the results after estimating equation (10) for gasoline prices. The results using diesel prices are similar. The marginal effect represents the average change in the market average price in response to a 10 percent increase in the independent variable. I find that if the number of firms increases by 10 percent, the average price will decrease by about 1 cent (in US dollars). If spatial differentiation increases by 10 percent, average price increases by about 1 cent. The effect is relatively big in terms of price margins which are about 10 cents per gallon. This result suggests that spatial differentiation allows stations to set higher prices. In addition, the estimated coefficients for the two variables on long-distance travelers suggest that prices are higher in markets with more long-distance travelers. The previous chapter shows a preference of stations to locate at different exits when there are more long-distance travelers, possibly in order to exploit the high stopping cost of travelers. The results in this chapter suggest that stations are indeed able to exploit these high stopping costs by setting higher prices.

5.2 Spatial Differentiation and Price Dispersion

This section describes the statistical relation between spatial differentiation and price dispersion. I use the coefficient of variation to measure price dispersion.[27] The coefficient of variation is the

[26] Not just entry stations.

[27] Other measures of price dispersion, like price range, standard deviation and variance provide similar results.

standard deviation divided by the average. I estimate the following equation:

$$\text{CoVar}_m = X_m'\beta + \gamma N_m + \delta MSpatDiff_m + \epsilon_m \tag{11}$$

where $\text{CoVar}_m = \frac{standard_deviation_of_price_m}{AvgPrice_m}$. X is a vector of market characteristics, N is the number of firms and $MSpatDiff$ is spatial differentiation. These are the same controls used to estimate average price.

Table 11 reports the results. I focus on the marginal effects that represent the average percentage change in the coefficient of variation in response to a 10 percent increase in the independent variable. I find that spatial differentiation is an important source of price dispersion. If spatial differentiation increases by 10 percent, price dispersion increases by around 10 percent for gasoline. The number of firms also increases price dispersion. Firms can be differentiated in other non-spatial characteristics, like brand, type of service, credit card policy, etc. Therefore the number of firms may capture non-spatial differentiation which increases price dispersion. On the other hand, the number of stations decreases price dispersion. Demand characteristics are statistically insignificant. It appears that only the degree of competition and product differentiation can explain price differences. At the same time there is significant variation left unexplained. This is consistent with Sorensen (2000). His intuition is that unobserved determinants of consumer search cause price dispersion to be higher. Long-distance travelers, in particular, have high search costs. Gasoline stations can exploit these high search costs and their own geographic location by setting high prices at different time periods each. This creates price dispersion within a given period. The result is also consistent with the theory that firms that are more differentiated can better price discriminate across locations, which also results in higher price dispersion.

6 Conclusion

Product differentiation is an important mechanism that firms use to gain market power. This paper shows empirical evidence of spatial differentiation of gasoline stations along interstate highways. The exogenous variation of the distances between interstate exits, allows me to study their effect on firms' location choices. Gasoline stations are attracted to places with higher demand, but when I control for local demand, the results suggest that stations establish at locations that are farther from other possible locations and therefore farther from potential competition. Firms differentiate in space in order to soften price competition and gain market power. In addition, I find that when entering the market, stations choose to locate farther from rivals in markets with greater competition, which confirms that there are incentives for spatial differentiation. Finally, I find that the average price and price dispersion are higher in markets with more differentiation. The former suggests that differentiation is a mechanism for gasoline retailers to gain market power, and the latter suggests that more differentiated firms can better price discriminate across locations.

The results are consistent with the theory that product differentiation is strong when consumers have strong preferences on locations,or when consumers have limited information about prices. In these cases, differentiated locations can provide significant market power even in cases where the product itself is homogeneous and price competition would otherwise have been severe. The results reject the theory that firms have a stronger incentive to move closer to competitors in order to "steal" customers. In the gasoline retail market, the product itself is homogeneous, therefore physical location is an important way for retailers to differentiate their product and gain market power.

This study cannot farther explain why spatial differentiation results in market power. We need to understand the demand side of the model in order to understand the sources of spatial differentiation. If consumer preferences on the highway are very strong, the incentives for stations to spatially differentiate will remain in place, as long as the preferences do not change. On the other hand, if long-distance travelers have incomplete information on prices on the highway, stations exploit their high search cost by spatially differentiating. In the latter case, a policy that gives incentives for stations to post prices on the highway can result in more severe price competition on the whole highway segment, and therefore the incentives to spatially differentiate will be lower. In addition, when the government constructs a new interstate highway, its decision on interstate exits can affect spatial competition and price competition in the later stages. This might be something that officials want to take into account.

An extension of this paper is to model the demand from long-distance travelers based on the traffic patterns, and model the price competition between profit maximizing firms. This will allow me to look at the sources of spatial differentiation and their effect on prices. For example, how much does the lack of price information affect prices?

References

Borenstein, S. and J. Netz (1999). Why do all the flights leave at 8 am?: Competition and departure- time differentiation in airline markets. *International Journal of Industrial Organization 17*(5), 611–640.

Donald, S. G. and K. Lang (2007). Inference with difference-in-differences and other panel data. *Review of Economics and Statistics 89*(2), 221–233.

Houde, J.-F. (2009). Spatial Differentiation and Vertical Contracts in Retail Markets for Gasoline.

Mazzeo, M. (2002). Product choice and oligopoly market structure. *Rand Journal of Economics 33*(2), 221–242.

Netz, J. and B. Taylor (2002). Maximum or minimun differentiation? location patterns of retail outlets. *Review of Economic and Statistics 84*(1), 162–175.

Pinske, J. and M. Slade (1998). Contracting in space. *Journal of Econometrics 85*(1), 125–154.

Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *Rand Journal of Economics 37*(3), 619–640.

Sorensen, A. (2000). Equilibrium price dispersion in retail markets for prescription drugs. *Journal of Political Economy 108*(4), 833–850.

Stavins, J. (1995). Model entry and exit in differentiated-product industry: The personal computer market. *Review of Economics and Statistics 77*(4), 571–584.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review 93*(2), 133–138.

Appendix

An Approximate Algorithm for the P-dispersion Problem

The discrete p-dispersion problem is the problem of selecting p out of n given points in some space, such that the distance sum between the p facilities is maximized. I need to assign stations such that they are maximally dispersed in the theoretical sense (equally spaced from each other), and I do that using an approximate method. In this appendix I outline the algorithm I use.

The algorithm has six main steps. For each market/line segment, I proceed as following:

1. Select the two extreme points along the line segment and define a normalized distance matrix between every pair of locations such that the distance between the two extremes is equal to one.

2. Define normalized positions from all points relatively to one of the extremes. Therefore one extreme will have a normalized position of zero, and another of one.

3. Define "fake" locations of stations such that they are maximally dispersed on a **continuous** line of length one. Therefore, I have the normalized actual positions of **interstate exits** from zero to one, and a second vector with "fake", theoretically maximum dispersed positions of **stations** from zero to one. The goal is to assign the stations from their "faked" positions to the normalized actual positions of exits.

4. Define a matrix with the distances between the actual normalized positions of exits and each "fake" point on the continuous line.

5. (a) First, assign the two extreme "fake" positions to the extreme actual positions. The first two stations are therefore easily being assigned to the to extreme exits.
   (b) Then, define a matrix with the minimum distances of "fake" positions to the exits and another matrix with the corresponding indices of exits.
   (c) Finally, assign stations/"fake" positions to exits sequentially, according to the smallest minimum distance: Assign each next station to the non-occupied exit that is closest to the "fake" position of the station. If the exit is occupied, move to step 6.

6. Assign the next station to the exit that is farther away from its closest occupied exit.

About 80 percent of the stations are assigned to exits at step 5, and the remaining at step 6.

Table 1: Descriptive Statistics - Interstate Exits

	Mean	Standard Deviation	Median	Min	Max
5,307 exits					
Gasoline stations	0.86	1.29	0.00	0.00	8.00
Traffic counts (average daily traffic)	31,345	23,746	24,250	2,995	212,025
Population of block groups within a 2 mile buffer zone	8,719	8,730	5,710	206	83,516
% of occupied houses that are rented	0.23	0.14	0.19	0.00	1.00
Median contract rent	389	175	356	0	2,001
Tax on gasoline, state + federal (cents per gallon)	39.5	5.4	38.4	25.9	52.4
Hotels and motels (zip code)	3.0	4.3	1.0	0.0	57.0
Restaurants (zip code)	7.0	8.4	4.0	0.0	67.0
	%				
Non-occupied exits (zero stations)	57.08				

Table 2: Descriptive Statistics - Markets

	Mean	Standard Deviation	Median	Min	Max
238 markets					
Interstate exits	22.3	12.4	19.0	4.0	59.0
Gasoline stations	19.2	11.9	17.0	1.0	58.0
Number of firms (number of brands + number of unbranded)	15.0	7.5	13.0	1.0	36.0
% of independent stations	0.40	0.18	0.38	0.00	1.00
% of major brand stations (Shell, BP, ExxonMobil, Chevron, Conoco)	0.24	0.16	0.24	0.00	0.80
Long-distance travelers (minimum traffic over locations)	21,790	13,418	18,850	2,365	76,730
Length of the segment/market (miles)	89.4	43.6	78.0	26.5	282.0

Table 3: Descriptive Statistics - Gasoline Stations

	Mean	Standard Deviation	Median	Min	Max
4,571 Stations					
Convenience Store	0.63	0.48			
% of Independently owned competitors	0.39	0.49			
% of major brand stations	0.27	0.44			
% of Shell stations	0.07	0.26			
Competitors in the market	25.53	12.43	25	0	57
Competitors at the same location	1.78	1.53	2	0	7
Distance from closest endpoint city (miles)	31.15	25.72	24.76	0.64	213.46

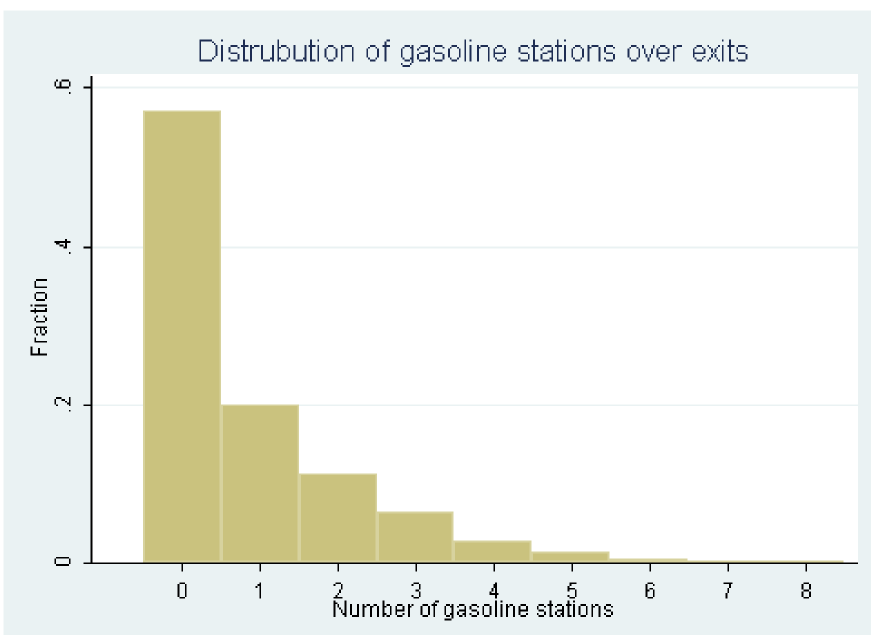

Figure 6: Distribution of gasoline stations over exits

Table 4: Effect of Geographic Isolation on the Number of Stations - by Adding More Demand Variables Each Time

	Specification	Coefficient of log(*Isolation2)*
1	Only measure of geographic isolation	-0.379***
		(0.0427)
2	Adding traffic, predicted long-distance travelers and dummy for important highway	-0.00129
		(0.0418)
3	Adding measures of local population (block group population around different ring zones)	0.103**
		(0.0430)
4	Adding entry cost variables, state gasoline taxes, number of hotels, restaurants, employees working in zip code, income and population of different census areas (tracts, zip codes, counties)	0.128***
		(0.0437)
5	Adding distance to nearest city and demographics of nearest city	0.127***
		(0.0436)
	Observations	5307
	Adjusted R-squared for specification 5	0.319

Robust, clustered (within markets) standard errors in parentheses
Market fixed effects included
Right hand side variables are in natural logarithm
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Effect of Geographic Proximity on the Number of Stations - OLS, Probit and 2SLS

	Specification	(1) OLS - number of stations	(2) OLS - number of stations at occupied exits	(3) Probit - entry at an exit	(4) 2SLS - number of stations
1	Only measure of geographic isolation	-0.379***	-0.0870	-0.484***	-0.480***
		(0.0427)	(0.0710)	(0.0791)	(0.0511)
…….					
3	Adding measures of local population (block group population around different ring zones)	0.103**	0.162***	0.076**	0.296**
		(0.0430)	(0.0615)	(0.0269)	(0.113)
…….					
5	Adding distance to nearest city and demographics of nearest city	0.127***	0.268***	0.081***	0.371***
		(0.0436)	(0.0681)	(0.0299)	(0.118)
	Observations	5307	2278	5301	5307
	Adjusted R-squared for specification 5	0.319	0.288		0.303

Robust, clustered (within markets) standard errors in parentheses

Market fixed effects in all regressions

Right hand side variables are in natural logarithm

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Effect of Geographic Proximity on the Number of Stations - Different Measures of Geographic Differentiation

		(3)	(1)	(2)
		Coefficient of log(*IsolationTWO*)	Coefficient of log(*Proximity*)	Coefficient of log(*Proximity - squared distance)*
1	Only measure of exit isolation/proximity	-0.379***	0.344***	0.186***
		(0.0427)	(0.0685)	(0.0240)
…..				
3	Adding measures of local population (block group population around different ring zones)	0.103**	-0.345***	-0.116***
		(0.0430)	(0.0618)	(0.0237)
…..				
5	Adding distance to nearest city and demographics of nearest city	0.127***	-0.416***	-0.137***
		(0.0436)	(0.0653)	(0.0243)
	Observations	5307	5307	5307
	Adjusted R-squared for specification 5	0.319	0.3175	0.312

Robust, clustered (within markets) standard errors in parentheses
Market fixed effects in all regressions
Right hand side variables are in natural logarithm.
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Least Squares Regression - Number of Stations At A Location

Specification 5	Location is a single exit	Location is a collection of nearby exits (within 1 mile)
Geographic proximity to other exits (log)	-0.401***	-0.870***
	(0.0658)	(0.112)
Selected demand characteristics		
Ratio of total traffic to long-distance travelers (log)	0.405***	1.157***
	(0.0759)	(0.121)
Predicted number of hwy drivers stopping for gas	0.0107	0.0284
	(0.0268)	(0.0522))
Exit to an important highway	0.217***	0.540***
	(0.0528)	(0.070)
Population of block groups within 2 miles (log)	0.381***	0.488***
	(0.0306)	(0.040)
Number of hotels in zip code	0.0175*	0.011
	(0.00898)	(0.014)
Distance to closest endpoint city (log)	0.0697*	0.216***
	(0.0440)	(0.068)
Entry cost variables		
Percentage of houses that are rented (log)	0.235***	0.472***
	(0.0331)	(0.055)
Median contract rent (log)	-0.0299	-0.089
	(0.0249)	(0.027)
Tax on gasoline (federal + state) (log)	-0.856*	-0.854
	(0.484)	(0.772)
Observations	5307	4220
Adjusted R-squared	0.3176	0.428

Robust, clustered (within markets) standard errors in parentheses

Market fixed effects included

*** p<0.01, ** p<0.05, * p<0.1

Table 8: Spatial Differentiation of New Entrants

	OLS - Sample restricted to entrants in the last two years	Marginal effect (10 % change)
Demand Characteristics		
Long-distance travelers (minimum traffic over locations)	0.0034***	0.87***
	(0.0009)	
Ratio of total traffic to long-distance travelers	0.0269	0.47
	(0.0231)	
Population of endpoint MSAs, 2007 ('00,000)	0.00080	0.00
	(0.0012)	
Competition variables		
Number of firms (number of brands + independents)	0.0088**	1.56**
	(0.0035)	
Number of stations	-0.0007	0.17
	(0.0019)	
Observations	180	
Adjusted R-squared	0.185	

Robust standard errors in parentheses.
Also included: exits per mile, segment length (miles), average exit geographic isolation and a constant.
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Descriptive Statistics - Market Price

	N	Mean	Standard Deviation	Min	Max
Stations					
Gasoline prices matched	90%				
Diesel prices matched	57%				
Markets					
Average market gasoline price	235	202.01	10.95	179.63	244.09
Average market diesel price	235	222.96	11.37	197.00	265.33
Coefficient of variation of gasoline price	235	0.030	0.016	0.000	0.097
Coefficient of variation of diesel price	235	0.032	0.021	0.000	0.140

Table 10: Average Market Price

	(1) gasoline price	Marginal effect (10% change) on cents $
Number of firms (number of brands + independent)	-0.725***	-1.09***
	(0.250)	
Number of stations	0.180	0.35
	(0.170)	
Spatial differentiation	10.36***	0.87***
	(3.969)	
Percentage of major-brand stations	20.02***	0.49***
	(4.887)	
Market size (selected characteristics)		
Long-distance travelers (minimum traffic over locations) ('000)	0.414**	0.90**
	(0.178)	
Average ratio of total traffic to long-distance travelers (minimum traffic)	2.808	0.42
	(1.829)	
Observations	235	
R-squared	0.305	

Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 11: Coefficient of Price Variation (log)

	(1) gasoline prices	Marginal effect (10% change) on % change
Number of firms (number of brands + independent)	0.0270* (0.0153)	4.06*
Number of stations	-0.0171* (0.00978)	-3.28*
Spatial differentiation	1.187*** (0.288)	10.02***
Percentage of major-brand stations	0.465* (0.248)	1.14*
Market size (selected characteristics)		
Long-distance travelers (minimum traffic over locations) ('000)	0.0043 (0.0073)	0.93
Average ratio of total traffic to long-distance travelers (minimum traffic)	-0.0646 (0.0716)	-0.96
Observations	231	
R-squared	0.298	

Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1