Spatial differentiation in retail markets for gasoline*

Jean-François Houde
University of Wisconsin-Madison

January 17, 2008

Abstract

This paper studies an empirical model of spatial competition. The main feature of my approach is to formally specify commuting paths as the "locations" of consumers in a Hotelling-type model of spatial competition. This modeling choice is motivated by the fact that consumers are moving across the market when consuming the product. Although this feature is perhaps more relevant for gasoline markets, this also applies to most retail markets since consumers are not immobile. The consequence of this behavior is that competition is not fully localized as in the standard address-model. In particular, the substitution patterns between stations depend in an intuitive way on the structure of the road network and the direction of traffic flows. Another feature of the model is that consumers' available options are directly linked with their commuting behavior; consumers who commute more encounter more stations and observe more prices. The demand-side of the model is estimated by combining a model of traffic allocation with econometric techniques used to estimate models of demand for differentiated products (Berry, Levinsohn and Pakes [1995]). The empirical distribution of commuters is computed with a shortest-path (or Dijkstra) algorithm, combining detailed data on the Québec City road network with aggregate Origin-Destination commuting probabilities. The model's parameters are then estimated using a unique panel data-set on the Québec City gasoline market from 1995 to 2001.
KEYWORDS: Spatial differentiation; Retail markets; Transportation; Market power.

1 Introduction

In this paper I estimate an empirical model of spatial competition. The main feature of my approach is to formally specify commuting paths as the "locations" of consumers in a Hotelling-type model of spatial competition. This modeling choice is motivated by the observation that consumers can

*First draft: September 2005. I am grateful for the advice and support of Chris Ferrall and Susumu Imai. I have benefited from the comments and discussions of Jason Allen, Robert Clark, Juan Esteban Caranza, Amit Ghandi, Beverly Lapham, Jeremy Lise, Sumon Majumdar, Jack Porter, Shannon Seitz, and seminar participants at Queen's, Concordia, UQAM, HEC-Montréal, Université de Montréal, University of Pennsylvania, Bank of Canada, Simon Fraser University, New York University, New York Fed, University of Wisconsin, the University of Northern-Illinois, ECSM 2006 in Minneapolis, SED 2008 in Prague, and the University of Chicago GSB. Funding from the Social Sciences and Humanities Research Council of Canada (SSHRC), and the *Fonds Québecois de la Recherche sur la Nature et les Technologies* (FCAR) are gratefully aknowledged. CORRESPONDENCE: (phone) (608) 262-3804, (email) houdejf@ssc.wisc.edu

be located at more than one point in the product space when deciding where to purchase gasoline. This characteristic is central to the analysis of gasoline markets, because consumers are by definition moving when consuming the product. Although this feature is perhaps most relevant for gasoline markets, it also applies to most retail markets since consumers are typically mobile when choosing where to go shopping. For instance, a grocery store located close to work or close home is potentially equally valuable for a consumer. Formally, I construct a new Hotelling-type address model which defines the locations of consumers as their home-to-work commuting path, rather than their home or workplace separately. The demand-side of the model is estimated using a unique panel dataset on the Québec City gasoline market between 1991 and 2001, following the techniques developed in Berry (1994) and Berry, Levinsohn and Pakes (BLP) (1995). The model is then used to measure retailer mark-ups, and quantify the market power of major retail chains.

Understanding the sources of market power in retail markets has important policy implications. Over the years, many provincial and state governments have attempted to limit the extent of market power of larger retailing chains. In gasoline markets, these concerns have led to the adoption of price floor regulations and contract restrictions.[1] These policies are typically aimed at protecting independent retailers from anti-competitive behavior by major chains. For instance, in 1997 the Québec provincial government established a price floor regulation after the occurrence of major price wars that spread throughout the Province. The design of the regulation is similar to an anti-dumping trade policy, and allows firms to sue their local competitors if they fix prices below a lower bound published every week by the Government.

A well specified model of demand is the key ingredient to evaluate the usefulness of these policies. In particular, locating consumers incorrectly in the product space could lead to biased elasticity of substitution between stores, and invalid predictions regarding firms' mark-ups in counter-factual experiments on market structure.

In gasoline markets, since consumers are not using the product at home, it is unclear why a station located anywhere along a common commuting path should be valued differently than a station close to home. To take this feature of the market into account, I define the location of consumers to be the set of intersections on a road network representing the shortest driving path

[1] In 1996, 12 US states and 4 Canadian provinces had price floor regulations targeted at gasoline retail prices, and 7 US states had contract restrictions preventing major Oil companies to be vertically integrated in the retail market (Petro-Canada 1996).

to commute from home to work (or school). By defining a location in this way, consumers choose where to buy gasoline by trading off price differences with the relative travel costs of deviating from their main commuting path. For a specific type of commuter, two gasoline stations at the two ends of the city can be close substitutes if they happen to be along his or her daily route. The estimated model therefore generates substitution patterns that are very different from the ones generated by a single location model.

Furthermore, the estimates of the model reproduce well-documented features of gasoline markets. For instance, Yatchew and No (2001) find that households who consume more gasoline also tend to pay lower prices, so that prices are endogenous in household demand. In the proposed model, long distance commuters naturally encounter more stations along their driving path and therefore may pay lower prices on average. On the other hand, a standard address model would have difficulty matching this fact without resorting to heterogeneous time costs, or reducing the cost of travel to account for distant gasoline purchases that would not be consistent with other consumer behavior.

In addition, contrary to the standard address model, competition is not fully localized in this framework since consumers can substitute stations far from each other but close to a commuting path. Consequently, even if consumers are not willing to deviate far from their path to shop for gasoline, price differences between two regions of a city are unlikely to be persistent if there is substantial commuting between the two regions. This is consistent with the observation that there is significantly less price dispersion within gasoline markets than in most other retail markets. For example the statistics reported in Lach (2002) suggest that the dispersion of gasoline prices is half as large as most grocery products. Gasoline price dispersion is similar in magnitude to the price of refrigerators, for which search costs are plausibly quite small relative to unit prices.

The estimation of the model is performed in two steps. First, I compute the empirical distribution of consumers across the road network using a deterministic route choice model, common in the transportation demand literature (Oppenheim (1995)). More specifically, the model predicts traffic on each segment of the road network, conditional on the distribution of origin-destination (OD) commuting pairs, assuming that individuals choose the shortest route in terms of time. The empirical distribution of commuters is obtained by combining the results of an OD survey conducted by the Québec Ministry of Transportation with aggregate population tables from the the Monthly

Labour Force Survey and the 1991, 1996, and 2001 Censuses of Canada. Optimal commuting paths are based on detailed data on the Québec City road network and a common algorithm used in the Geographic Information System (GIS) literature.

Conditional on the empirical distribution of commuting paths, the preference parameters of the model are estimated using Generalized Method of Moments (GMM), following the techniques developed by Berry (1994), Berry, Levinsohn and Pakes (1995), and Nevo (2001). In particular, the estimation methodology deals explicitly with the endogeneity of prices using an instrumental variable approach, controlling for time invariant unobserved gasoline station heterogeneity.

The estimation of the demand parameters is performed using a panel of gasoline stations in the Québec City area between $1995 - 2001$. During this period important changes were underway in the structure of the North-American gasoline retail industry, associated with massive exit of stations and entry of new categories of retailers (e.g. large stations with a convenience stores). Also, for more than half of the sample periods, the market is subject a price floor regulation. These two characteristics of the sample present important sources of exogenous variation in the data, changing substantially the choice set of consumers over time, and reducing the correlation of prices with respect to unobserved station characteristics.

The results can be summarized as follows. First, the model based on commuting behavior is shown to fit the observed distribution of sales more closely than the standard home-address model. In particular, stations with high market shares are not the ones located close to home. This leads to a correlation close to zero (or even negative) between the number of people living in a neighborhood of stations and their sales. The estimated model also reveals important differences between the home-location model and the commuting location model. The estimated transportation cost is negative under the home-location model, leading to negative shopping cost (i.e. consumers value positively the time necessary to shop for gasoline). The commuting model on the other hand provides an estimate of the shopping cost which is high and provides a realistic estimate of consumers' value of shopping time.

The strength of market power under the two models is also different. The estimates from the home-location model predict that the cross-price elasticities between products should increase in distance because of the negative transportation cost. Also, holding parameters constant across the two models, the degree of competition between stores is much more localized with the home location model than with the commuting model. In particular, the cross-price elasticities fall sharply with distance in the traditional model. These differences translate into different estimates of firms' markups. In particular, profit margins are smaller and less dispersed in the commuting model. This confirms that firms have less local market power when consumers have multi-dimensional locations.

Finally, the overall degree of market power is estimated to be quite small despite the presence of large retail chains. The role of independent stations is found to be marginal in this market, because they are on average much less efficient than stations selling national brands (i.e. low quality products and higher costs). The results suggest that keeping an artificially high number of independent stations in the market, through a price floor regulation for instance, can result in a higher equilibrium price.

My paper is related to a large literature in empirical industrial organization devoted to the estimation of discrete choice models of demand using firm level data (see Bresnahan (1987) for an early example, and Berry (1994) and Berry et al. (1995) for further developments of the estimation methodology). Recently, the methodology proposed by Berry et al. has been extended in several directions to evaluate market responses to policy changes. Examples of this approach include the evaluation of international trade policies (e.g. Berry et al. (1999) and Brambilla (2004)), the analysis of mergers (e.g. Nevo (2000) and Dubé (2005)), and the valuation of new goods (e.g. Petrin (2002)). See Ackerberg, Benkard, Berry and Pakes (2005) for an extensive review of this literature.

Recently researchers have studied empirically markets for spatially differentiated products. Closely related to the techniques used here, Davis (2006) and Manuszak (2001) extend the BLP methodology to estimate an address-model applied respectively to the US movie theater industry and two Hawaiian gasoline markets. Thomadsen (2004) also estimates an address model using data on prices and store characteristics in the fast-food industry, imposing the equilibrium conditions of a Bertrand pricing game with spatial differentiation[2]. While my approach shares econometric

[2] Other papers looking empirically at markets with spatial differentiation include: Smith (2004) studies demand for grocery products using micro-data from a UK household survey, Pinkse, Slade and Brett (2002) and Pinkse and Slade (2004) estimate a model of spatial competition applied respectively to the wholesale gasoline and beer pints markets, the papers by Hastings (2004) and Hastings and Gilbert (2002) use a natural experiment (namely an observed vertical merger in California) to measure market power and the impact of vertical integration in markets for gasoline, and McManus (2006) estimate a model of spatial differentiation in the market for coffee shops in order to study the extent of non-linear pricing.

techniques with these papers it differs on the demand side of the model. Specifically, I consider the possibility that consumers have multiple locations through their commuting behavior. The theoretical literature on spatial competition has extended the Hotelling model in many directions, including recently the possibility of multiple dimensions of differentiation on the firm side (see Anderson, de Palma and Thisse (1992) for a review of this literature). The impact of consumers' commuting behavior on oligopolistic price competition has been studied theoretically in Claycombe (1991) and Raith (1996). Both papers show that under certain conditions markets with commuting consumers are more competitive than the traditional Hotelling model. The results presented in this paper confirm that this is indeed the case in gasoline markets.

The rest of the paper is organized as follows. The next section presents the data. Section 3 provides evidence of the importance of commuting patterns in explaining demand for gasoline. Section 4 present a structural demand model, and Section 5 discusses the estimation and identification strategy. Sections 6 presents the empirical results, including an evaluation of market power conducted using the estimated parameters. I conclude the paper and discuss extensions in section 7. Further computational and data construction details are placed in the Appendices.

2 Data overview

2.1 Gasoline retailing data

The gasoline station data used to estimate the model has been collected by Kent Marketing, the leading survey company for the Canadian gasoline market. The panel spans 41 bimonthly periods (every 2 months) between 1995 and 2001 for every station in the Qébec City market. The survey offers very accurate measures of sales and station characteristics since each site is physically visited at the end of the survey period, and volume sold is measured by reading the pumps meters[3]

The observed station characteristics include the type of convenience store, a car-repair shop indicator, the number of service islands, opening hours, type of service, and an indicator for the availability of a car-wash. A major brand indicator is also added to the set of characteristics to reflect the fact that consumers might view major gasoline brands as higher quality. The major

[3]Note that approximately 10% of the sample, firms refused to participate in the survey for some or every periods. For those observations, the station characteristics (including prices) are accurately measured, but the volume sold is not available. Since the inversion procedure cannot accommodate missing values in the quantity variable, I imputed the missing values using linear regression methods. The explanatory variables include the average neighborhood market shares, a polynomial function of the geographic coordinates of the locations, prices, characteristics and lagged sales (for stations who were previously participating in the survey).

Table 1: Description of Station Characteristics

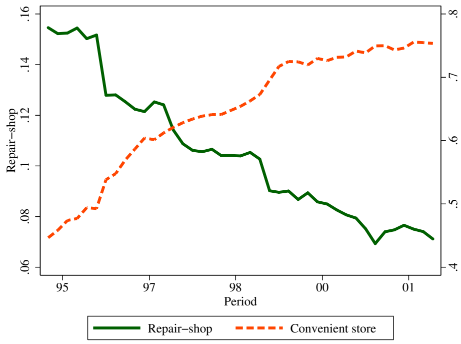
	Mean	Std-Dev.	Min	Max
Convenient Store Size	1.4	1.3	0	3
Repair Shop	.17	.37	0	1
Number of Islands	2.3	1.4	1	8
Opening Hours Category (24 hours = 1)	.4	.49	0	1
Type of Service (Self-Service =1)	.62	.48	0	1
Carwash	.2	.4	0	1
Major Brand	.64	.48	0	1

retailers include five chains who are integrated in the refinery sector: Shell, Esso/Imperial Oil, Ultramar, Irving, and Petro-Canada. The sample includes 12,477 observations, for 429 different gasoline station sites. On average each product is observed for 33 periods.

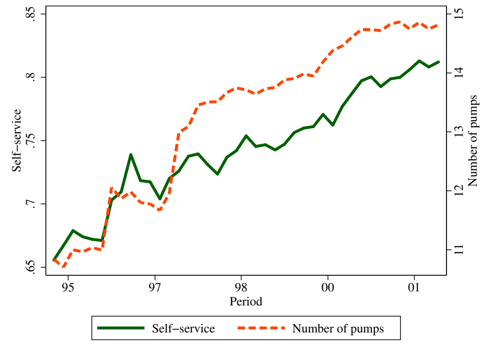
Over the period studied, the North-American gasoline retail industry underwent a major reorganization, associated with massive exit and entry of new categories of stations (see Eckert and West (2005)). These changes were mainly due to technological innovations common to most retailing sectors which increased the efficient size of stations (e.g. automization of the service, better inventory control systems), as well as changes in the preferences of consumers for certain amenities (e.g. decreased use of small repair garages). Figure 1 summarizes these trends in the Québec Cit market. Figures 2(a) confirm that the proportion of stations with a convenience store, and the proportion of self-service stations have increased by roughly 30% and 24% respectively. At the same time, the proportion of stations offering car-repair services dropped by 15%. These changes have been induced by the exit of a large number of small-capacity stations and the entry of large capacity stations, as presented in Figure 2(b) shows. Figure 2(c) shows that the number of stations also dropped by close to 20%. Note that it is mainly the major chains who are responsible for these changes, which explains their larger market share at the end of the period. As shown in Figure 2(c), the majors[4] have steadily increased their market share, without expanding their retail network. Figure 2(d) also presents two important facts about this market; the small cross-sectional dispersion in prices, and the low markups. Over most of the period, the inter-quantile range of prices (i.e. difference between the 75th and the 25th quantile) was smaller than 1¢/L, or about

[4] Throughout the text, the "majors" will refer to the five retail chains who are vertically integrated in the refinery sector; Ultramar, Esso (or Imperial), Shell, Petro-Canada, Irving. Jointly they controlled about 80% of the market shares, but only about 60% of stations in Québec City.

Figure 1: Illustration of the key market trends

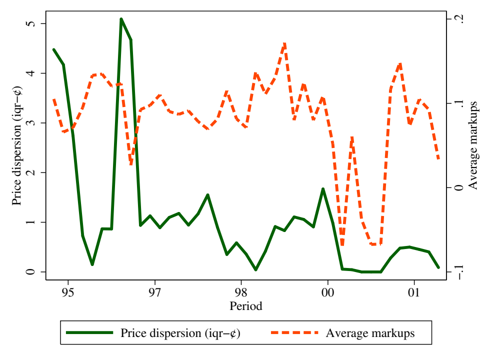


(a) Repair shop and convenient stores



(b) Capacity and type of service

(c) Number of stations

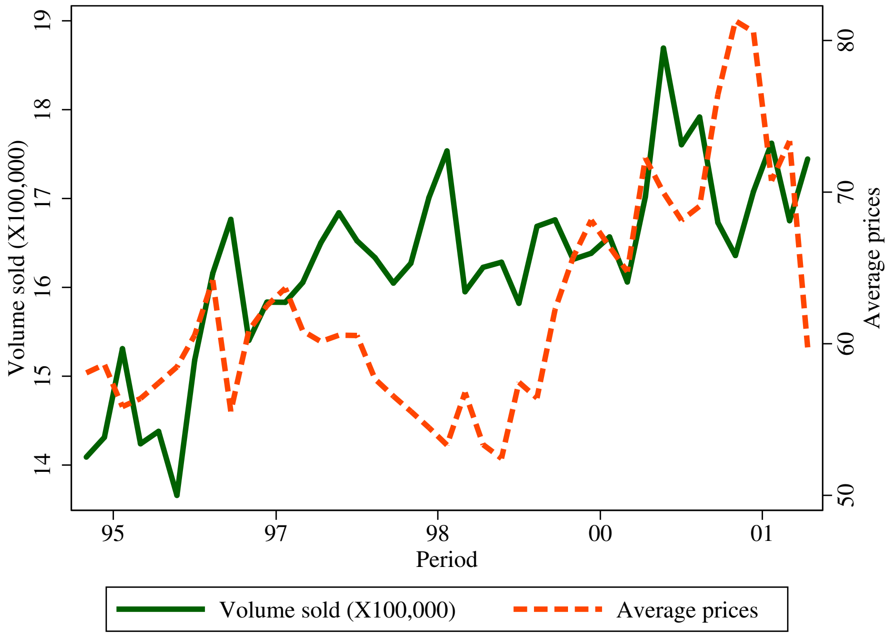


(d) Price dispersion and markups

1.5% of the average price. The markups, calculated as one minus the ratio of the rack price to the average price, oscillated 8% or 9% for most of the period. Finally, Figure 2 shows that both aggregate prices and demand are very cyclical, and have been steadily increasing over the 1990s. These upward trends reflect the improvements in economic conditions and the city population growth, and the sharp increase in the world price of oil after 1999.

Another important characteristic of the Québec City market is the presence of a price floor regulation which aims at protecting small independent retailers against presumed predatory pricing strategies by major chains. The *law on petroleum products* was created in the summer of 1997 and administrated by the Régie de l'énergie du Québec (hereafter the board). This followed a major price war during the summer of 1996, and over part of the year 1997. The mandate of the board

Figure 2: Evolution of Demand and Prices in Québec City



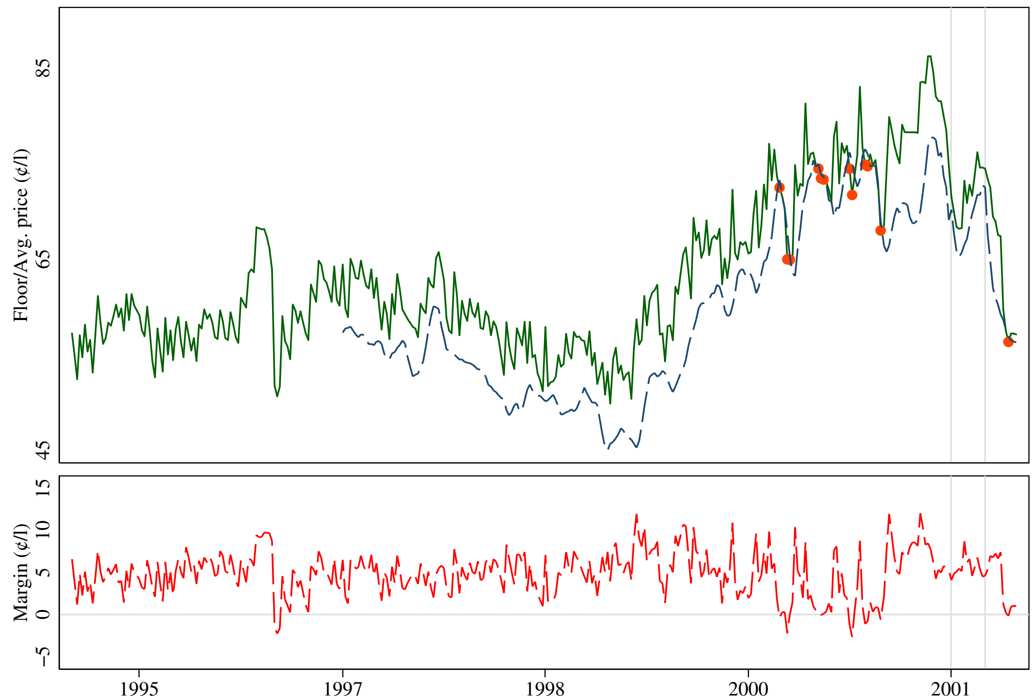
is to determine a weekly floor price (or Minimum Estimated Price (MEP)) and prevent price wars by imposing a minimum margin if necessary in a designed geographic market.

The MEP is the sum of the minimum rack price published every week by refiners, taxes and an estimate of the transportation cost from the terminal to each station. The role of the MEP is to set a floor price under which a firm can sue its competitor(s) to be financially compensated on the basis of excessive and unreasonable commercial practice. This new feature of the law facilitates suing procedures between competitors in the market, in a similar fashion as anti-dumping laws. The other part of the regulation, and the one that has been discussed more heavily, is the ability of the board to impose an additional minimum margin to the MEP. It was added to the regulation in August 1999. It allows the board to add 3¢/L to the calculation of the MEP in a specific region after the occurrence of a period of low enough prices.[5] Since 1999 this minimum margin regulation has been imposed three times, including in Quebec City in the fall of 2001.

Figure 3 presents the evolution of the weekly average price in the city, as well as the price floor

[5]The determination of this minimum margin followed form a calculation of the average operating cost of a representative station in the province. After public consultations, the board has decided that the representative station is a self-service station operating a convenient store, and having an annual sale volume of 350 million liters.

Figure 3: Evolution of the average weekly price, price floor and margin



Vertical grid lines = added 3¢/l minimum margin ; Red dots = binding price floor.

and the average retail margin (i.e. price - rack price). The red dots identify weeks when the average market price was equal to or smaller than the price floor, and the two vertical gray lines indicate the imposition of the additional minimum margin. The first thing to note from the figure is that the price floor regulation does not appear to be binding before 2000, or before the addition of the minimum margin regulation. In fact, apart from an episode in 1996 and after 2000, the average retail margins has been oscillating around 4¢/L. Another characteristic of the price of gasoline is its high short-run volatility. From the figure we see that retail margins seem to be following a two or three weeks cycle, as documented, among others, by Noel (2005).

2.2 Construction of the empirical distribution of commuters

In the absence of micro-data surveying jointly consumers' commuting and gasoline purchasing decisions, we rely on market-level data on commuting and gasoline consumption. The empirical method proposed here is to use the empirical distribution of commuters to predict demand of gasoline at the store level. Before discussing the details of the store-choice model, I first describe the construction of the empirical distribution of commuters. I then use the predicted distribution of traffic flows to evaluate empirically the importance of commuting in predicting demand for gasoline.

The geography of the market, here defined as the Québec city Census Metropolitan Area (CMA), is described by a grid of L location areas where people reside and/or work, and a road network $G = (N, A)$ where N denotes a set of street intersections (or nodes) and A is the set of road segments (or arcs). The construction of the distribution of commuters across the road network involves two elements: (1) the empirical distribution of people over origin and destination locations, and (2) the route used to commute between those points.

Consumers are characterized by a pair $(s, d) \in \mathcal{L}^2$ of origin and destination locations. I consider two types of consumers: local and outside commuters. The location pairs for local commuters correspond to the centroid of their area of residence and the location of their main occupation (i.e. work or study). Outside commuters on the other hand are assumed to travel along the main highways of the city, and therefore each outside commuter origin/destination locations correspond to the beginning and end of a particular highway segment (I consider 11 such segments in the empirical analysis).

The distribution of consumers in period t across origin-destination locations is given by T^t_{sd}. Local commuters' occupations can either be working (W), full-time studying (S) or working at home (U). The total number of type (s, d) consumers is thus given by:

$$T^t_{sd} = W_{it}\Omega^{W_t}_{sd} + S_{it}\Omega^{S_t}_{sd} + U_{it}\mathcal{I}(s = d) + O_t\mathcal{I}((s, d) \in \mathcal{H}), \quad (1)$$

where $\Omega^{k_t}_{sd}$ is the probability of commuting between locations (s, d) for occupation k, O_t is the number of outside commuters and $\mathcal{H}$ is the set of locations characterizing each highway segment. Since commuting probabilities are only available for local commuters I use the average number of occupied hotel room in the CMA to approximate the quantity of outside commuters. I further assume that each of those travelers is equally likely to be "located" along one of the 11 segments of the highway network of Quebec city. The number of outside commuters O_t is thus the total number of outside commuters divided by the number of highway segments.

Commuting probabilities are computed from a pair of surveys conducted in 1996 and 2001 by the Québec government in the Québec City CMA. The results of the survey are available in the form of aggregate OD tables, providing the predicted number of commuters between every pair of Traffic

Table 2: Descriptive Statistics of the Location Areas

	Mean	Sd-Dev	Q25	Q50	Q75
Population	1089	1310	385.5	513.8	1076
Workers	629.8	743	227.6	311.3	639.5
Students	117.5	164.8	40.52	62.57	108.4
Area size (km^2)	6.296	28.47	0.1809	0.4991	2.536
Number of areas	501				
Number of OD pairs	36,402				

Area Zones (TAZs).[6] The tables are available for four categories of trips and transportation modes and three periods of the day. I use the results of two OD tables for every transportation mode, representing work and study trips over a full day window.[7] Each OD matrix is then converted into a commuting probability matrix by dividing each row by the total number of trips originating from each TAZ. In order to predict traffic in between the two survey dates, the OD probabilities are interpolated assuming a constant growth rate.[8]

The empirical distribution of local commuters living in each origin location s is computed by combining data from the three most recent censuses (i.e. 1991, 1996 and 2001) and the monthly Canadian Labour Force survey (available at the CMA level) to predict the month-to-month variations in the relevant population measures. Table 2 presents a set of statistics describing the distribution of population across the 501 location areas used in the empirical analysis. Note that the distribution of area sizes is highly skewed because the Quebec CMA includes the rural fringes of the city. The median location area is quite small however, measuring $0.5km^2$ and populated by 513 individuals.

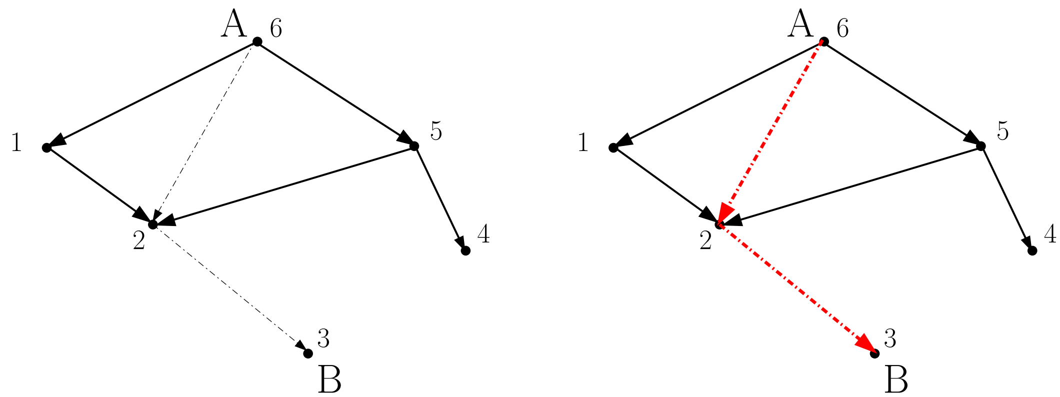
In order to predict traffic along each road segment I assume that commuters choose the optimal route between (s, d) by minimizing the travel time between their home and main occupation locations. This assumption corresponds to the deterministic route choice model used to predict traffic

[6]The aggregate OD matrices are freely available on the Ministry website: http://www1.mtq.gouv.qc.ca/fr/services/documentation/statistiques/enquetes/index.asp. The survey report available on the same website provides further details on the conduct of the survey and the method used to aggregate individual responses (MTQ 2002).

[7]Adding leisure and shopping trips is conceptually feasible, but would increase significantly the computation cost of the model. In addition, since consumers are commuting to their workplaces on a daily basis, it is the most appropriate commuting path to characterize preferences for gasoline station locations.

[8]The OD probabilities were further disaggregated into a finer grid in order to predict commuting patterns more accurately. Appendix A describes in detail the method used to compute the Ω^k_{sd} for each location pairs.

Figure 4: Example of Shortest Paths on a Small Network



(a) Network Representation (b) Shortest Path from A to B

Source	Dest.	Distance (c_{sd})
1	2	$\sqrt{8}$
2	3	$\sqrt{10}$
5	2	$\sqrt{10}$
5	4	$\sqrt{8}$
6	1	$\sqrt{15}$
6	2	$\sqrt{5}$
6	5	$\sqrt{5}$

(c) Node and Arc Relations

patterns in the transportation literature (see Oppenheim (1995)). It generates a single path for each type of consumer, denoted by $r(s, d)$, which abstracts from any congestion or unobserved preferences considerations.[9] This optimal route is calculated using a version of the Dijkstra's Shortest Path Tree (SPT) algorithm (Shekhar and Chawla (2003)). Figure 4 illustrates the calculation of a shortest path between a pair of locations (A, B) on a fictitious network. The procedure iterates on the travel time between the starting node and every other nodes on the network. In the example, the optimal path is given by $r(A, B) = \{6, 2, 3\}$, and the travel time is given by $t(r(A, B)) = \sqrt{5} + \sqrt{10}$.

In practice, I compute 501 shortest-path trees recording the shortest paths from the centroid of each location area to every nodes in N, as well as the travel times and distances. Information on the Quebec city CMA street network was obtained from the CanMap RouteLogistics database

[9]The no-congestion assumption is realistic in the Québec City area, since the population is spread over a large territory and the road network is well developed. It has been used also by Thériault et al (1999) to study the distribution of commuting trips in the Québec City metropolitan area using similar data.

Table 3: Description of the Québec City Road Network

Number of nodes	23, 394
Number of arcs	32, 167
Avg. length (meters)	217, 8
Avg. travel time (minutes)	0.28

(DMTI-Spatial 2004), the leading road data provider in Canada. Table 3 describes the size of the road network data. Note that the street/intersection data is extremely fine. It includes more than 30, 000 street segments, and the average travel time is less than 30 seconds.[10]

3 Preliminary evidence on the importance of commuting

Before turning to the structural model, it is valuable to compare the empirical distribution of commuters described above with the distribution of gasoline station market shares. The objective of this exercise is to provide a first test for which consumer location model (i.e. home-address vs commuting-address) is more appropriate to explain the spatial distribution of demand for gasoline.

I do so first by regressing the observed market share of stations on the number of individuals living and commuting within three Euclidean distance buffers around each station (i.e. $B_1 = [0, 1/3)$, $B_2 = [1/3, 1/2)$ and $B_3 = [1/2, 1)$). Under the assumption that consumers shop for gasoline only around their home location, the home-buffers measure the relevant local market size of stores. The traffic-buffers on the other hand measure the relevant market size under the assumption that the location of consumers is their home-to-work driving path. In that case for instance the number of consumers in the first buffer (i.e. $[0, 1/3)$) corresponds to the number of commuters such that the smallest Euclidean deviation from their path is less than 300 meters.

The results are summarized in Table 4. From this table we see that the market share of stations is negatively correlated with the number of people living close-by. In fact the coefficients on three home buffers increase with the distance bands, suggesting that consumers are not shopping for gasoline close to their home. This indicates that it will be difficult to reconcile the data with a single address model in which consumers have positive transportation costs.

The second and third specifications reveal that gasoline stations sales are positively and signifi-

[10]More details on the computation of the shortest paths are provided in Appendix B.

Table 4: Regression of observed station market shares on the size of three home and commuting distance buffers.

	Home-market	Traffic-market	Both-markets
	(1)	(2)	(3)
Commuters in: $d \leq 300m$		.0607***	.0633***
		(.0024)	(.0025)
Commuters in: $300m < d \leq 500m$		.0015	.0074***
		(.0015)	(.0016)
Commuters in: $500m < d \leq 1km$		.0089***	.0199***
		(.0017)	(.0019)
Residents in: $d \leq 300m$	-.0244***		-.0288***
	(.0016)		(.0016)
Residents in: $300m < d \leq 500m$	-.0141***		-.0213***
	(.0016)		(.0017)
Residents in: $500m < d \leq 1km$	.0167***		-.0044**
	(.0017)		(.0019)
Obs.	12477	12477	12477
R^2	.0269	.0911	.1151

The explanatory variables are expressed relative to their standard deviations. Robust standard-errors are in parenthesis.

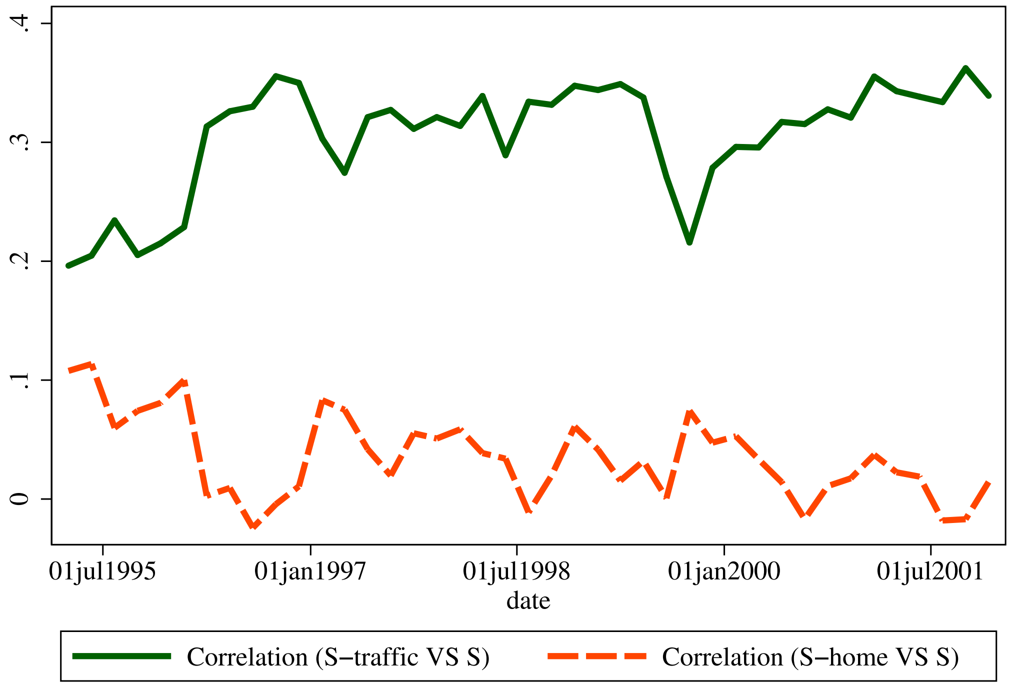
cantly correlated with the number of commuters in a small neighborhood. Moreover, this correlation is falling with distance from stations, suggesting that consumers incur large and positive costs of deviating far from their home-to-work commuting paths.

A second way of comparing the distribution of consumers with stations' market shares is to use a simple model in which consumers randomly choose a store within a certain distance band around their home and/or commuting path. Formally, if $\bar{d}^h$ and $\bar{d}^c$ denote the maximum Euclidian distance that a consumer is willing to travel to buy gasoline away from home, the market share of station j is given by:

$$S_j^h = \sum_{s,d} \mathcal{I}(d(s,j) < \bar{d}^h) \frac{1}{\sum_j \mathcal{I}(d(s,j)<\bar{d}^h)} T_{s,d},$$
$$S_j^c = \sum_{s,d} \mathcal{I}(d_{min}(r(s,d),j) < \bar{d}^c) \frac{1}{\sum_j \mathcal{I}(d_{min}(r(s,d),j)<\bar{d}^c)} T_{s,d},$$

where $d(s,j)$ is the Euclidian distance from home location s to station j, and $d_{min}(r(s,d),j)$ is the smallest Euclidian distance from the set of nodes in driving path $r(s,d)$ to station j. To evaluate empirically the predictions of these two models the value of $\bar{d}^h$ and $\bar{d}^c$ are calibrated to maximize the average period-by-period correlations between the observed market shares and the predicted

Figure 5: Correlations between observed market shares and predicted market shares from the home and commuting buffer models.



Note: Market shares from Traffic and Home are computed with a buffer of 300m and 1km repsectively.

ones from the two buffer models.

Figure 5 presents the correlation coefficients for the 42 sample periods of data. The first striking feature of this figure is that the correlations generated from the simple commuting buffer model are systematically larger than the ones generated from the home buffer model. In fact, most of the correlations between the data and the predictions of the home buffer model are very close to zero, while the ones from the commuting model all lie around 0.3.

Moreover, the size of the distance bands that maximize the correlations are estimated to be slightly more than 300 meters for the commuting model, and 1.10 kilometers for the home model. Since these parameters measure the importance of transportation cost in this simple buffer model, this suggests that the single address model will systematically under-estimate the value of consumer transportation costs relative to a model which take into account consumer mobility.

The previous two exercises looked at the spatial correlation between gasoline sales and the distribution of commuters/residents. Another way of comparing the two definitions of consumer location is to study how the entry or exit of a store affects adjacent stores. To do this, I consider three definitions of neighborhood:

1. Immediate neighbors: $B^1_{jt} = \Big\{ k \in J_t \Big| d^t_{jk} < b_1 = 2 \text{ minutes} \Big\}$,
2. Fraction of common potential consumers:

$$B^2_{jt} = \left\{ k \in J_t \left| \frac{\sum_i \mathcal{I}(d_{min}(r_i, j) < \bar{d}) \times \mathcal{I}(d_{min}(r_i, k) < \bar{d})}{\sum_i \mathcal{I}(d_{min}(r_i, j) < \bar{d})} \leq b_2 = 50\% \right. \right\},$$

3. Common street neighbors: $B^3_{jt} = \Big\{ k \in J_t \Big| \text{Street}_j \bigcap \text{Street}_k \neq \emptyset \Big\}$,

where J_t is the set of active stations in period t, d^t_{jk} is the travel time between station j and k, $\mathcal{I}(\cdot)$ is an indicator function, and $\bar{d} = 300$ meters is the maximum deviation from commuting path. Using these definitions, I calculate the number of exits and entrants in each neighborhood B^k_{jt}. Table 5 present the results of two linear regressions of the change in firm j market share on the number of entrants and exits in those neighborhoods.[11]

The first specification uses a count of the number of exits in each neighborhood between period t and $t-1$ as explanatory variables, while the second specification uses the number of new stations in each neighborhood. The results reveal no significant correlation between the number of new entrants in all three neighborhoods and the change in market shares. This is likely due to the fact that new entrants tend to locate themselves in neighborhoods where no incumbent station is already active, leaving little variation in the explanatory variables. However, the first specification reveals that the number of exits in the "common consumers" and "common street" neighborhoods are significantly and positively related with changes in market shares. This suggests that substitution patterns between stores are not as localized as what the home-location model would predict. According to the single-address model, consumers should be substituting toward immediate neighboring stations after the exit of a store, which would generate a positive and significant correlation between the number of exits in the first neighborhood and changes in market shares. The fact that the coefficients on the two last neighborhood definitions are significant suggests instead that consumers are substituting

[11] The depend variable is measured as the log of the ratio of station j share on the share of the outside option. In the multinomial model of demand, this corresponds to the mean quality of product j (see Berry (1994)). Defined in this way, the dependent variable measure the relative change in market shares controlling for changes in the market size.

Table 5: Regression of the change in stations' market shares on the number of exits and entrants in three neighborhoods

	Exit	Entry
	(1)	(2)
Travel distance ($d < 2$ minutes)	.012	-.001
	(.010)	(.013)
Common consumers ($d > 50\%$)	.021	.001
	(.008)***	(.015)
Common street	.025	.015
	(.007)***	(.017)
Obs.	11738	11738

The dependent variable is defined as: $\Delta\delta_{jt} = \log\left(q_{jt}/(M_t - \sum_k q_{kt})\right) - \log\left(q_{jt-1}/(M_{t-1} - \sum_k q_{kt-1})\right)$, where M_t is the market size (in liters) and q_{jt} is the average daily sales at station j in period t (in liters). Robust standard errors are parenthesis.

toward stores located along their commuting paths and/or connected by a common street, but not necessarily physically closed in terms of travel distance. Of course these results must be interpreted with caution given the endogeneity of exit decisions. In particular, if the sales of stations located the same immediate neighborhood are affected by a common unobserved characteristic that also affects their decision to exit, the coefficient on the number of exits in the travel distance neighborhood will be biased toward zero.

In sum, these results reveal that the distribution of commuters is significantly correlated with the distribution of gasoline sales in the market, while the distribution of the population is not. High market share stations are therefore not located close to dense neighborhoods, but rather at the intersection of large commuting paths. The strength of the correlation between the traffic buffer and the observed market shares also validates the assumptions made to construct the empirical distribution of commuters. In particular, it is reassuring that the traffic-buffer model explains a relatively large fraction of the observed variance in shares without any additional station characteristics controls, and despite the restrictive assumptions that consumers only buy gasoline on their way to work and that route choice is fully deterministic. Jointly these correlations highlight the importance consumers' commuting behavior in the determination of demand for gasoline, and in the estimation of consumers' transportation costs.

A structural model of demand for gasoline is needed, however, to understand the role of commuting in shaping the substitution patterns between products and measure the degree of market power. The previous analysis lacks most importantly two key elements to measure accurately the degree of spatial differentiation: the valuation observed and unobserved station characteristics, and the availability of an outside option. Ignoring the first element significantly lower the predictive power of the buffer measures, since two neighboring stations are treated as fully homogenous products. Furthermore, the previous analysis neglects the fact that consumers are not purchasing gasoline with equal probability. For instance, consumers living in a high density neighborhood, where a larger fraction of individuals are unemployed or work close to home, are unlikely to buy gasoline every week. The number of residents and commuters in a certain distance band will overestimate the potential market size of stations located in those neighborhoods, thereby reducing the explanatory power of these variables. The structural model is also useful to translate the demand estimate into measures of profits and cross-price elasticities. The next two sections present a model and estimation strategy to accomplish these objectives.

4 Model

I model demand for gasoline as a discrete choice problem over $J+1$ options.[12] In particular, a consumer of type $r(s,d)$ has the option of buying gasoline from one of J stores or use an alternative mode of transportation (i.e. option 0). The indirect utility of buying option j conditional on commuting along path $r(s,d)$ is given by:

$$u_{ij}\big(r(s,d)\big) = \begin{cases} X_j\beta - \alpha p_j - \lambda_1 D\big(r(s,d),j\big) + \xi_j + \epsilon_{ij} & \text{if } j \neq 0, \\ \lambda_0 t(s,d)) + \epsilon_{ij} & \text{otherwise,} \end{cases} \tag{2}$$

where X_j is a vector of observed station characteristics, p_j is the average posted price, $D\big(r(s,d),j\big)$ is the shortest Euclidian distance from driving path $r(s,d)$ to station j, ξ_j is an index of unobserved (to the econometrician) station attributes, $t(s,d)$ is the home-to-work commuting time computed using the shortest-path algorithm described above, and ϵ_{ij} is an *iid* random utility shock distributed according a type-1 extreme value distribution.

The set of time varying station characteristics includes the number of gas pumps, the number of service islands, the type of service, the type of convenience store (if any), dummy variables indicating whether the station offers car-repair and/or car-wash services, an indicator for brand, and a set of

[12]Time subscripts are omitted in this section to reduce the notational burden.

time-dummy variables capturing unobserved period-specific variables (e.g. weather, price of public transportation). Since these variables mainly characterize the type of amenities offered by station j, the unobserved attribute ξ_j measures the set of characteristics of the location which are valued positively by all consumers. These include, for instance, time invariant characteristics of the streets or neighboring stores (e.g. side of the street, size of the parking lot, road congestion), as well as time varying factors affecting the value of the location (e.g. road construction).

As noted by Petrin (2002) and others, the inclusion of the utility shock, ϵ_{ij}, in the model can generate unrealistic substitution patterns across products because of the independence of irrelevant alternative assumption imbedded into it. In particular, without other sources of heterogeneity between consumers, the cross-elasticities of substitution depends solely on the relative valuation of products. To get around this problem, the payoff function described in equation 2 introduces two sources of heterogeneity: (i) the location of consumers with respect to stations, and (ii) the valuation of the outside option. The first component insures that consumers will substitute toward products that are close to each other, given the distance metric $D\big(r(s,d),j)\big)$ defined in section 2.2. For instance, as long as λ_1 is positive and large relative to the other coefficients, consumers will unlikely deviate far from their commuting paths to buy gasoline. The model will therefore predict large cross-elasticity of substitution between stations that are physically close to each other and/or connected by popular commuting paths. The other source of consumer heterogeneity is related to willingness to pay for the outside option. In particular, if $\lambda_0 > 0$, consumers with long home-to-work commutes are more likely to buy gasoline than consumers who live close to their workplace. If this is confirmed in the estimation, this would reflect the fact that the value of using an alternative mode of transportation (e.g. public transportation or car pooling) is decreasing in the travel time between (s,d).

In the model consumers are also heterogeneous with respect to their transportation needs per day, denoted by $\bar{q}\big(r(s,d)\big) = c_0 + c_1 m(s,d)$, where $m(s,d)$ measures the length of path $r(s,d)$ in kilometers.[13] This representation implies that individual consumption is split between heterogeneous work commutes and a common fixed quantity c_0, representing leisure and shopping trips. In

[13] An alternative representation of the decision could allow consumers to make a mixed continuous and discrete choice over how much and where to buy gasoline. For instance Smith (2004) and Berkowitz et al. (1990) study empirically both the intensive and extensive margin of the decision with micro data on household consumption of grocery products and gasoline respectively, using a model derived from Dubin and McFadden (1984) I chose to use pure discrete choice model here because of the nature of the data available (i.e. aggregate sales at the station level).

the empirical analysis, the value of c_1 will be fixed to the average gasoline consumption of a car in a city (i.e. 0.12 liters/km) and c_0 will be estimated.

Given the distribution assumption on ϵ_{ij}, the conditional probability of buying from store j, for a consumer commuting along route $r(s,d)$ takes the familiar multinomial logit form:

$$P_{j|r(s,d)} = \frac{\exp\Big(\delta_j - \lambda_1 D\big(r(s,d),j\big)\Big)}{\exp\big(\lambda_0 t(s,d)\big) + \sum_k \exp\Big(\delta_k - \lambda_1 D\big(r(s,d),k\big)\Big)}, \tag{3}$$

where $\delta_j = X_j\beta - \alpha p_j + \xi_j$ is the mean value of store j.

The predicted demand at the station level is then obtained by aggregating individual choice probabilities over every OD pair:

$$Q_j(\delta) = \sum_{s,d} \bar{q}\big(r(s,d)\big) P_{j|r(s,d)}(\delta) T_{sd}. \tag{4}$$

5 Estimation methodology

The set of preference parameters to be estimated are given by $\theta = \{\alpha, c_0, \lambda_0, \lambda_1, \beta\}$, where α is the price sensitivity parameter, λ_0 is the travel cost for the outside alternative, λ_1 is the parameter entering the transportation cost function, and β is the vector of parameters entering the station characteristics value equation. The main dataset used is an unbalanced panel of observed sales and product characteristics: $Y_t = \Big\{\big\{q_{jt}, p_{jt}, X_{jt}\big\}_{j\in J_t}\Big\}$, $t = 1..T$.

The methodology used to estimate the model builds on the techniques developed by Berry (1994) and Berry, Levinsohn and Pakes (1995) to estimate discrete choice models of demand using aggregate data. More specifically, since δ is a linear function of product characteristics and prices, a non-linear GMM estimator can be used to estimate the model by dealing with the endogeneity of prices with respect to the structural error without assuming a parametric distribution function on ξ_{jt}.[14]

An endogeneity problem arises for two reasons. First, gasoline prices are known to adjust very frequently, on a weekly or even daily basis (see for instance Noel (2005)). This introduces a

[14]The computation cost of the inversion step necessary to compute the empirical moments is rapidly increasing in the number of markets (or time periods) and products. In order to increase the speed of convergence, I use a Newton-Raphson (or Broyden's) root-finding algorithm instead of the contraction mapping proposed by Berry et al., and paralleled the task so that each processor is inverting the demand system for a subset of the sample periods. Appendix C.1 describes the algorithm used to invert the system of market share equations.

measurement error in prices if the adjustment process affects the position of a station in the period-distribution of prices. For instance, the observed sales of a station might have been generated by a sequence of low price that is unmeasured given that I use a weighted average of the beginning and end of period prices to compute p_{jt}. Secondly, since firms and consumers observe the quality index ξ_{jt} when making their decisions, prices will adjust in the short-run to changes in the unobserved product quality. To control for this correlation, I decompose the unobserved quality into a permanent and transitory component: $\xi_{jt} = \bar{\xi}_j + \Delta\xi_{jt}$. As discussed above the fixed component $\bar{\xi}_j$ refers to characteristics of the location associated with the road network and the organization of the city. These include, for instance, how easy it is to enter the parking lot of a store, and which side of the street the station is located. The second component is associated with temporary changes made to the quality of the location, such as employee turnover or temporary road repair.

I use two sets of moment conditions to identify the model.[15] Following Nevo (2001), the first set of moment conditions combines an Instrumental-Variable (IV) approach with fixed-effects at the station's location level. If $\tilde{w}_{jt} = w_{jt} - \frac{1}{n_j}\sum_t w_{jt}$ denotes the within$-j$ transformation of a variable w_{jt}, this first set of empirical moment conditions are given by:

$$\bar{g}^1_n(\theta) = \frac{1}{n}\sum_{j,t} g^1_{jt}(\theta) = \frac{1}{n}\sum_{j,t} \Delta\xi_{j,t}(\theta)\tilde{W}^1_{jt}, \tag{5}$$

where n is the number of observations, and W^1_{jt} is a vector of predetermined variables including k characteristics in X_{jt} and l instrumental variables Z_{jt}.

I follow the suggestion of Berry et al. (1995) to construct instruments. They propose a set of IVs which measure the firm's own and rivals' product characteristics. These variables are valid instruments since they enter naturally the equilibrium pricing rule in any Bertrand game with product differentiation. In addition, if firms choose their product characteristics after learning about the value of their own and their neighbors' quality shocks, these IVs are independent of ξ_j. I use a similar strategy here by constructing IVs which measure the average stations' neighborhood characteristics.[16] To capture the idea that consumers can substitute toward products that are not

[15] Manuszak (2001) uses a different identification strategy to estimate a similar model of spatial competition in gasoline markets. In particular, he imposes the Nash equilibrium condition on prices, as in the original application of Berry, Levinsohn and Pakes (1995). I chose to use a different identification strategy to avoid the misspecification bias induced by imposing a potentially invalid pricing rule. It is likely to be the case in my application because of the price regulation, and the fact that retail markets for gasoline are characterized by alternating periods of price wars and (tacit) collusion.

[16] Davis (2006) and Manuszak (2001) constructed similar instruments based on the fact that competition is highly localized in most spatial differentiation models.

necessarily physically close to each other, I use two definitions of a station neighborhood. The first is based on the physical distance between products (i.e. driving time), and the second reflects the degree of connectivity between stations (i.e. common streets). Let $d(i,j)$ denotes driving time in minutes between location i and j, and $S(i.j)$ denotes an indicator function equal to one if i and j share one street. Two sets of instruments for each observed characteristics k can therefore be constructed:

$$\begin{aligned}\bar{X}_{jtk}(a,b) &= \frac{1}{n_{jt}(a,b)}\sum_{j'} X_{j'tk}\mathcal{I}\big(a \leq d(j,j') < b\big) \\ \bar{X}^{s}_{jtk} &= \frac{1}{n^{s}_{jt}}\sum_{j'} X_{j'tk}\mathcal{I}\big(S(j,j')=1\big),\end{aligned}$$

where $n_{jt}(a,b)$ and n^s_{jt} are the number of active stations in driving time neighborhood (a,b) and along a common street, respectively. I consider three driving distance rings in the empirical application: $(a,b) \in \big\{[0,1/3),[1/3,1),[1,2)\big\}$. The vector instrumental variables also includes the number of active stations in each neighborhood (i.e. $n_{jt}(a,b)$ and n^s_{jt}).

Note that the presence of time-invariant unobserved product characteristics can reduce the validity of these instruments in this context. In particular, the exogeneity assumption of the station characteristics X_{jt} and the IVs, Z_{jt},with respect to the structural error ξ_{jt} is unrealistic if $\bar{\xi}_j$ is shared by nearby stations. If stations endogenously choose their location based on these common locations' unobserved attributes, the instruments measuring the characteristics of neighborhood competitors will be correlated with ξ_{jt}. The fixed-effects approach gets around this problem by assuming instead that the transitory component of the unobserved quality of products is exogenous with respect to the inter-temporal variation of station characteristics $\tilde{X}_{jt}$ and instruments $\tilde{Z}_{jt}$. The identification assumption is therefore that stations observe the current transitory quality shocks after entering or changing their observed characteristics. This is reasonable given the lumpy nature of the characteristics (e.g. convenient store, capacity, etc.).

This approach exploits the richness of the time variation induced by the entry, exit and reconfiguration of stations over the sample period. These changes provide a valid source of exogenous variation in the instruments since they are related to technology innovations not specific to the Québec City market. Without such inter-temporal variation in the market structure, the within transformation of the variables would eliminate all variation in the instruments. Moreover, adding geographic markets in the analysis, as in Nevo (2001), would not introduce additional variation.

This is because the product fixed effects are by definition specific to each geographic market, contrary to other product differentiation applications in which we observe the same product in different markets.[17]

The second set of moment conditions is obtained by matching the proportion of people who use their car to go to work (or study) in the model with the empirical frequencies from the Origin-Destination survey conducted in Québec City in 2001. In particular, letting q^0_{sd} denote the observed proportion of car users for the pair of origin and destination (s, d), I compute the following moment conditions:

$$\bar{g}^2_{n_2}(\theta) = \frac{1}{n_2} \sum_{s,d \in \text{Workers}} (P_{0|sd}(\theta) - q^0_{sd}) W^2_{sd}, \tag{6}$$

where W^2_{sd} includes a constant and the commuting time for a consumer of type (s, d) (in logarithm of commuting hours). Note that this specification of the micro-moments is equivalent to an indirect inference approach, which would force the model to replicate the observed linear reduced-form relationship between commuting time and the log car-usage probability. In the 2001 OD survey, the OLS estimation of this reduced-form relationship gives the following results:

$$(1 - P_{r0}) = \underset{(0.0121)}{0.9645} + \underset{(0.00484)}{0.1159} \times \log(t(s, d)), \quad R^2 = 0.15 \tag{7}$$

which indicates that consumers living far from their workplace are significantly more likely to use their car (standard errors are in parenthesis).

Joining equations 5 and 6, the GMM objective function is given by:

$$\begin{aligned} Q(\theta) &= \bar{g}_n(\theta)\Phi^{-1}\bar{g}'_n(\theta), \quad \Phi = V_n^{-1} \\ \text{where} \quad & \bar{g}_n = [\bar{g}^1_{n_1} : \bar{g}^2_{n_2}]. \end{aligned} \tag{8}$$

The weighting matrix obtained using a two-step procedure which takes into account the spatial and time-correlation between observations, following the approach suggested by Conley (1999). Appendix C.3 describes in further details the construction of the the weighting matrix.

Finally, estimation and inference on parameters is conducted using the Laplace-type estimator proposed by Chernozhukov and Hong (2003). The advantages of this estimator in this context are twofold. First, in the current application, the objective function of the non-linear GMM problem

[17] Nevo's application was the ready-to-eat cereal market, for which he observes the same brands in a cross section of cities over time.

exhibits multiple local optima and flat regions, which significantly complicates the numerical solution of the problem. The Laplace-type estimator, which is based on an MCMC algorithm, does not suffer from this problem since it is a global optimization method. Secondly, Chernozhukov and Hong (2003) show that as the number of MCMC simulation draws goes to infinity, the mean and standard-deviation of the posterior distribution of θ corresponds to its asymptotic distribution counterpart. Moreover, the posterior distribution of any function $f(\theta)$ of the parameters also corresponds to the asymptotic distribution. This means that estimation and inference can be done simultaneously for all parameters of interest, including composite parameters like the average markups of store-level price elasticities. Appendix C.2 presents in more details the MCMC algorithm used.

5.1 Identification

In this section I discuss how the moment conditions presented above identify the key parameters of the model, that is the price coefficient (α), transportation cost (λ_1), disutility of commuting (λ_0), and baseline consumption (c_0).

The parameters α and λ_1 determine how much a consumer is willing to deviate from his path to save on price. Jointly they are thus responsible for shaping the substitution patterns between stores. As discussed above, the level of the price coefficient is identified from the orthogonality assumption of the instruments with respect to the inter-temporal change in the quality of a station. The instruments measure the strength of local competition, and therefore affects the cross-sectional distribution of prices without being correlated with changes in the location quality index $\Delta\xi_{jt}$.

How the first set of moment conditions identifies the transportation cost is less obvious. As discussed in section 4, if the data are generated by a model with positive and large transportation cost relative to the value of δ_{jt}, the cross-elasticities of substitution will depend on the distance between products and the "connectivity" of locations. The first set of moment conditions puts restrictions on these substitution patterns by forcing the change in the quality of stations to be independent of changes in the composition of local markets. To see this, consider the exit of a store which positively affects the market share of close-by stations and change value of the instrumental variables. If the estimation algorithm assigns a low value on the transportation cost, the model will predict a raise in the quality of continuing stations in order to match the observed raise in the market share of continuing stations, which in turns will violate the independence assumption between the instruments and $\Delta\xi_{jt}$. If instead the transportation cost is high, the model will be able to explain this market share increase without affecting the quality levels, therefore satisfying the moment conditions. The previous discussion therefore suggests that the first set of moment conditions separately identifies the price coefficient and the transportation cost, provided that there is enough inter-temporal variation in the structure of local markets.

Finally, as in Petrin (2002), the second set of moment conditions directly identifies the two remaining non-linear parameters. In particular, the average proportion of work-commuters identifies the baseline consumption level c_0, and the observed positive relationship between the proportion of car users and commuting distance identifies the disutility of commuting λ_0.

6 Results

GMM estimates of the model parameters are presented in Table 6 for the multi-address (or commuting) model, and in Table 7 for the standard home address model. The estimated parameters, standard-errors and confidence intervals are computed using $30,000$ replications of the MCMC algorithm. Figure 6 presents the simulated Markov chains of four parameters from the first specification. From these pictures, we see that the markov process is stationary, and therefore that the distribution of the simulated parameters corresponds to the pseudo-posterior distribution.

Two sets of instruments were used to estimate the models. Specifications 1 and 3 use a small number of IVs, measuring only the average capacity and the number of neighboring stations. Specifications 2 and 4, on the other hand, use an extensive set of characteristics describing neighboring stations as in Berry et al. (1995). A drawback of using a large number of neighboring stations' characteristics is that some are not highly correlated with prices, leading to a weak instruments problem. For instance, in a regression of prices on stations' own characteristics, time dummies and neighbors characteristics, the F-test corresponding to the joint validity of the instruments is equal to 6.69 in the case of the large IV set and 18.02 for the small IV set. While both tests are significant at all levels, the first one suggests a weak correlation between prices and some of the competing stations' characteristics.

The results from the commuting model show that the price and transportation cost coefficients are significantly smaller with the restricted set of instruments than with the larger set. Table 6 also reports the ratio of the transportation cost to the price coefficient, which measures the price

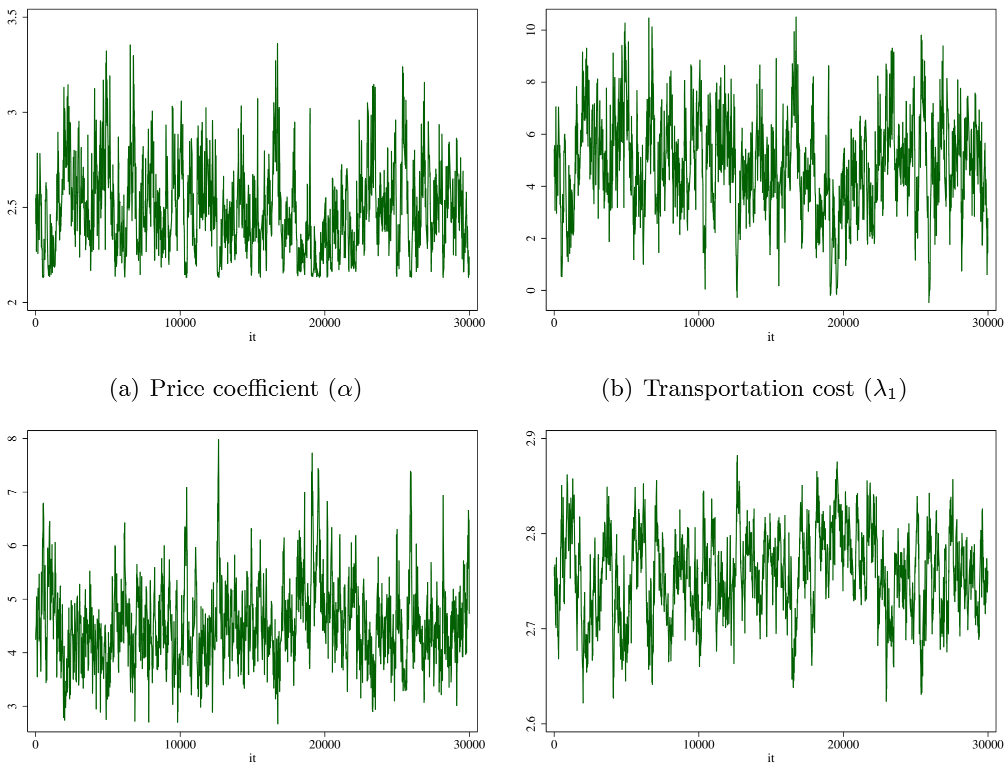
Table 6: GMM estimation results for the multi-address (commuting) model

	(1)			(2)		
	Param. (s.e.)	95% CI		Param. (s.e.)	95% CI	
Baseline consumption – c_0	2.75 (0.0454)	(2.67	2.84)	2.81 (0.0294)	(2.76	2.87)
Commuting time (×100 hours) – λ_0	4.31 (0.7760)	(3.23	6.35)	5.79 (0.4852)	(4.96	6.86)
Price (×10¢/l) – α	2.51 (0.2307)	(2.14	3.01)	1.92 (0.0363)	(1.86	2.00)
Transportation cost (×10km) – λ_1	5.11 (1.8633)	(1.11	8.46)	1.75 (0.4813)	(0.77	2.64)
Shopping cost – λ_1/α	2.04 (0.5802)	(0.52	2.81)	0.91 (0.2339)	(0.41	1.32)
Avg. price elasticity	−14.842 (0.9017)	(−16.677	−13.256)	−11.447 (0.1724)	(−11.821	−11.022)
# Pumps	0.01 (0.0003)	(0.01	0.01)	0.01 (0.0002)	(0.01	0.01)
Islands	0.04 (0.0021)	(0.04	0.04)	0.04 (0.0007)	(0.04	0.04)
Mixed service	−0.04 (0.0045)	(−0.05	−0.03)	−0.06 (0.0010)	(−0.07	−0.06)
Self service	0.06 (0.0117)	(0.03	0.08)	0.06 (0.0027)	(0.06	0.07)
Small conv. store	−0.04 (0.0117)	(−0.06	−0.02)	−0.03 (0.0008)	(−0.03	−0.03)
Medium conv. store	−0.03 (0.0105)	(−0.05	−0.01)	−0.02 (0.0009)	(−0.02	−0.02)
Large conv. store	0.07 (0.0128)	(0.05	0.09)	0.08 (0.0009)	(0.07	0.08)
No repair-shop	0.20 (0.0335)	(0.12	0.25)	0.12 (0.0112)	(0.10	0.14)
Carwash	−0.07 (0.0014)	(−0.07	−0.07)	−0.06 (0.0006)	(−0.06	−0.06)
Extended hours	−0.47 (0.0176)	(−0.49	−0.42)	−0.48 (0.0075)	(−0.49	−0.46)
24 hours	0.08 (0.0178)	(0.05	0.12)	0.03 (0.0038)	(0.03	0.04)
Time dummies	included			included		
Brand dummies	included			included		
Location fixed effects	included			included		
Nb. observations	12,477			12,477		
Nb. moments	10			42		
Objective function	30.872			149.156		

Table 7: GMM estimation results for the single-address (home) model

	(3)			(4)		
	Param. (s.e.)	95% CI		Param. (s.e.)	95% CI	
Baseline consumption – c_0	2.73 (0.0803)	(2.55	2.87)	2.80 (0.0361)	(2.73	2.87)
Commuting time ($\times$100 hours) – λ_0	3.08 (2.4247)	($-$0.01	9.88)	4.86 (0.9031)	(3.25	6.79)
Price ($\times$10¢/l) – α	2.27 (0.0974)	(2.02	2.41)	1.99 (0.0466)	(1.88	2.07)
Transportation cost ($\times$10km) λ_1	$-$2.53 (1.5541)	($-$4.49	1.94)	$-$1.44 (0.4744)	($-$2.26	$-$0.37)
Shopping cost – (λ_1/α)	$-$1.11 (0.6826)	($-$1.86	0.97)	$-$0.72 (0.2252)	($-$1.09	$-$0.20)
Avg. price-elasticity	$-$13.650 (0.1125)	($-$13.875	$-$13.425)	$-$11.892 (0.1616)	($-$12.2152	$-$11.569)
# Pumps	0.01 (0.0001)	(0.01	0.01)	0.01 (0.0001)	(0.01	0.01)
Islands	0.03 (0.0068)	(0.02	0.05)	0.04 (0.0017)	(0.03	0.04)
Mixed service	$-$0.04 (0.0065)	($-$0.04	$-$0.02)	$-$0.07 (0.0014)	($-$0.07	$-$0.06)
Self service	0.08 (0.0035)	(0.07	0.09)	0.07 (0.0002)	(0.07	0.07)
Small conv. store	$-$0.01 (0.0068)	($-$0.03	$-$0.00)	$-$0.03 (0.0023)	($-$0.03	$-$0.02)
Medium conv. store	$-$0.01 (0.0022)	($-$0.01	$-$0.01)	$-$0.02 (0.0005)	($-$0.03	$-$0.02)
Large conv. store	0.09 (0.0109)	(0.06	0.10)	0.08 (0.0036)	(0.07	0.08)
No repair-shop	0.14 (0.0199)	(0.08	0.16)	0.11 (0.0091)	(0.09	0.13)
Carwash	$-$0.10 (0.0164)	($-$0.12	$-$0.06)	$-$0.08 (0.0062)	($-$0.09	$-$0.06)
Extended hours	$-$0.41 (0.0155)	($-$0.42	$-$0.36)	$-$0.47 (0.0065)	($-$0.48	$-$0.46)
24 hours	0.06 (0.0057)	(0.05	0.07)	0.03 (0.0033)	(0.02	0.04)
Time dummies	included			included		
Brand dummies	included			included		
Location fixed effects	included			included		
Nb. observations	12,477			12,477		
Nb. moments	10			42		
Objective function	41.077			147.085		

Figure 6: Simulated Markov chains for four of the main parameters



(a) Price coefficient (α)

(b) Transportation cost (λ_1)

(c) Disutility of commuting (λ_0)

(d) Baseline consumption (c_0)

difference necessary to make a consumer indifferent between a station located directly on its path and a station located 1 kilometer away. This statistic is estimated to be 2.04¢/L in specification 1, and 0.91¢/L in specification 2. To transform these values into an estimate of the value of shopping time, we need to decompose the cost of shopping 1 kilometer away from an ideal location as a time cost and a gasoline consumption cost:

$$\text{Shopping cost}(\Delta d = 1\text{km}) = \underbrace{\text{time-value} \times 2\text{km} \times \frac{1}{50}\text{hours/km}}_{\text{Time cost}} + \underbrace{p \times Q \times 2\text{km} \times 0.1\text{l/km}}_{\text{Gasoline cost}}. \tag{9}$$

Evaluating the above expression for a consumer purchasing 30 liters of gasoline at a price of 60¢/L, the two transportation cost estimates correspond to a value of time of 12$/hour in specification 1 and 3.83$/hour in specification 2. Specification 1 therefore offers a more realistic portrait of consumer behavior. However, both values can be considered to be high given the level of price dispersion in the market. According to Figure 1, the difference between the 75th and 25th quantile of the price distribution was almost always around 1¢/l during the period. The estimates show that if consumers require a price difference somewhere between 1¢/L and 2¢/L to deviate 1 kilometer away from their location, they will very likely consume gasoline along their driving path.

On the other hand, the results of the home-location model suggest that consumers have a negative or zero transportation cost. While the price coefficient is similar across all four specifications, the point estimate of the λ_1 is negative in specifications 3 and 4, and not statistically different from zero in specification 3. This translates into a negative ratio λ_1/α, suggesting that consumers prefer to buy gasoline far from their home location and have a negative value of shopping time. This result is consistent with the negative or zero correlation between the distribution of gasoline sales and the population distribution documented in section 3, and confirms that the single-address model is not appropriate in explaining gasoline shopping decisions.

The rest of the parameters are qualitatively similar across specifications. The baseline consumption is estimated to be around 2.75 liters per day, which corresponds to an average weekly consumption of 28.75 liters for a worker commuting 15 kilometers every day. The disutility of commuting (λ_0) is positive, implying that long commuters are more likely to buy gasoline than short commuters.

The value of stations' characteristics variables are also intuitive. In particular, consumers tend to prefer larger stations (both in terms of the number of service islands and pumps), self-service over full or mixed service, stations with large convenient stores, and stations opened 24 hours per day. On the other hand, the average consumer values negatively car-wash facilities and repair-shops. Overall, these results suggest that consumers prefer amenities that increase the rapidity and availability of the product (e.g. self-service or 24 hours dummies), and that reduce congestion at the station (e.g. number of pumps and islands).

6.1 Evaluation of market power

In this section I use the estimated parameters to analyze the level of market power, and compare the predictions of the two location assumptions. The key difference between the two spatial models resides in the shape of the predicted substitution patterns across products. To compare these, I compute the cross-price elasticities between products using the parameters from Specification 1, and vary the location of consumers. The objective of this exercise is to evaluate how the cross-price elasticities change with the physical distance between products. Table 8 presents two sets of results

Table 8: OLS regression of cross price elasticities on two measures of distance

	Commuting Model	Home Model
Travel time (hours)	-1.125	-1.731
	(0.008)	0.013
Share of common commuters	3.318	5.757
	(0.063)	0.100)
Intercept	1.223	1.317
	(0.005)	0.008)
Observations	189,727	189,727
R^2	0.18	0.21

Robust standard errors in parentheses. Dependent variables are expressed relative to their period averages. Sample = 10 percent random sample cross-elasticity pairs.

regressing a 10% random sample of cross-price elasticities on the travel time between two stores and the proportion of common commuters. The results show that the cross-price elasticities fall rapidly with the distance between stores, and increase in the proportion of common commuters. This relationship is much stronger, however, for the home location model; a one percent decrease in the travel time raise the cross-elasticity by 3.11% for the home location model, and by only 1.44% for the commuting model. A similar relationship is found with respect to the proportion of common commuters. If we measure the strength of competition between two stores by their cross-price elasticities, the home location model clearly predicts that competition is more "localized" than the commuting model. According to the home-location model stations compete mainly with their immediate neighbors, which leads to more local market power than what the commuting model predicts. Therefore, everything else being equal, a market characterized by consumers located at a single point is less competitive than a multi-address market.

Next, I measure market power following Nevo (2001), and recover an estimate of station-level marginal costs from the first-order conditions of a static pricing game. In particular, if $\mathcal{J}_f$ represents the set of stations owned by firm f, a Bertrand-Nash equilibrium is characterized by the following J non-linear equations:

$$Q_j(p) + \sum_{k \in \mathcal{J}_f} (p_j - c_j) \frac{\partial Q_k(p)}{\partial p_j} = 0, \tag{10}$$

where c_j is the constant marginal cost of store j. Let $\Delta(p) = \left[\frac{\partial Q_k(p)}{\partial p_j}\right]$ be the $J \times J$ matrix of cross and own price derivatives, and Ω be a $J \times J$ matrix describing the structure of the market (i.e. $\Omega_{ij} = 1$ if station j and i jointly set their price). For a given assumption on Ω, the vector of marginal costs can be recovered from the observed prices p_t as follows:

$$c_t = p_t - \Big[\Omega * \Delta(p_t)\Big]^{-1} Q(p_t). \tag{11}$$

Since the exact ownership contracts between chains and stations are not observed, any assumption on Ω is somewhat arbitrary. There is however some evidence that chains behave as multi-product oligopolists, even if they do not fully own every station in their network. For instance, all five of the vertically integrated chains monitor prices constantly in most urban markets of the Province, and advise their stations on the timing and level of price changes. Also, Ultramar, the leading chain in the market, advertises a "lowest-price guarantee" policy in the neighborhood of all Utramar stations. This implicitly constrains the price of individual stations, whether or not they are owned by the chain. For these reasons, in what follows I assume that the observed prices are generated by a static Bertrand pricing game with multi-product oligopolists (i.e. chain-pricing). Moreover, rather than modeling the price floor regulation explicitly, I consider only the periods for which the constraint does not appear to be binding (i.e. $1995 - 1999$).

Table 9 presents the estimated and counter-factual markups assuming that prices are generated by the chain-pricing structure (i.e. column 2). The first two lines of the table compare the average and standard-deviation of markups for three alternative market structures: (i) store level pricing, (ii) chain pricing, (iii) collusion. The first thing to note from this table is that markups are significantly smaller than what previous researchers have found in other industries with differentiated products. For instance, Nevo (2001) reports median price-to-cost margins closer to 46% in the ready-to-eat cereal market, and Berry et al. (1995) reports car markups that are in a 15% to 30% range. This suggests that the level of market power is very small in gasoline retail markets. The estimates are also similar to the average markups computed using the posted rack price in Figure 2(d) (i.e. between 8% and 10%). The key factor explaining this low average markup is the large estimated price coefficient which leads to a store-level average price elasticity around -15 (see Table 6).

Moreover, average markups are low in the market despite the fact that prices are set at the chain level, which is equivalent to a tacit collusion between stations of the same chain. Recall that these chains jointly controlled close to 80% of the market in 2001. The first column of Table 9 shows that if prices were instead set at the station level, the average markups would decrease

Table 9: Counter-factual markups under two locational assumptions

	Market Structure Store	Chain	Collusion	Δ% (Chain-Store)	Δ% (Collusion-Chain)
Multi-address model:					
Markup dispersion	0.0273	0.0587	0.0465	54.9579	−26.1460
Avg. markups	0.0742	0.0803	0.1430	7.6571	43.7627
Single-address model:					
Markup dispersion	0.0377	0.0643	0.0548	42.4703	−20.0598
Avg. markups	0.0752	0.0813	0.1440	7.4111	43.4886
Δ% (Home-Commuting) Markup dispersion	29.0533	8.5460	12.7868		
Δ% (Home-Commuting) Avg. markups	1.4182	1.1555	0.6783		

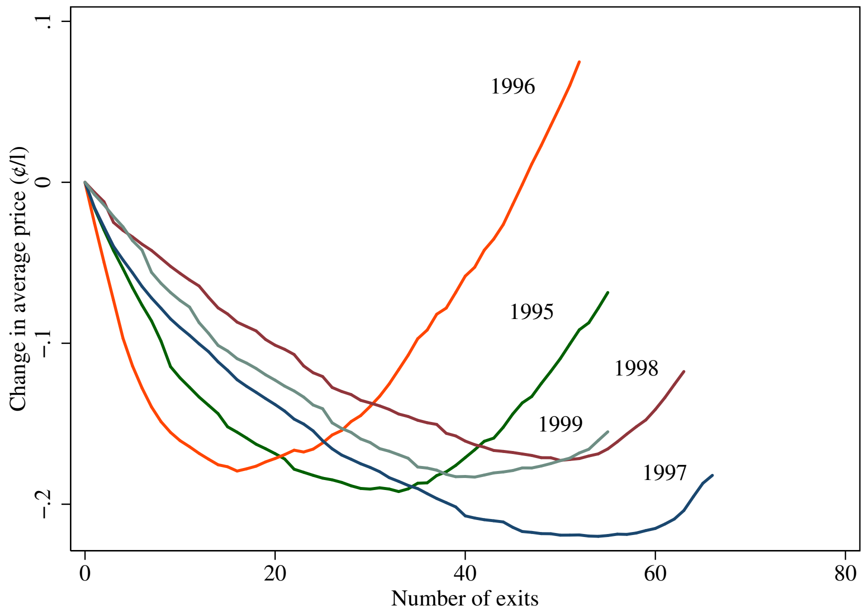
Markup dispersion is measured as the standard-deviation of log markups. Each entry is generated from the Nash equilibrium prices corresponding to each specific market structure, using the parameters from Specification 1. The marginal costs and mean valuations are recovered assuming multi-product price competition and a specific location structure (i.e. home or commuting addresses).

by only 0.6 percentage point. This result, combined with the counter-factual markups under full collusion (third column), shows that the presence of large retail chains falls short of realizing the collusive profits. To see this, note that while markups increase by 7.65% because of the presence of multi-product firms, markups would increase by 44% if all stations were colluding.

The fourth row of Table 9 presents the predicted markups from the home-location model. The comparison between the two models on this margin is difficult because the quality index δ_j and the marginal cost must be different in order to match the observed market shares and prices. The reported numbers are the markups predicted by the home-location model using the parameters estimated in Specification 1, since the transportation cost estimate is negative in specifications 3 and 4. As expected, for all three market structures the markup differences between the two models are systematically positive in favor of the home-location model. Prices and markups are predicted to be about 1% higher by the home location model compare to the commuting model. The main reason for the small magnitude is the fact that consumers are very price sensitive, which means that the market is very competitive whether or not consumers have more than one location.

One notable difference between the two models however is in the dispersion of markups, which measures the ability of firms to price discriminate spatially across consumers. An interesting feature

Figure 7: Change in average market price following the counter-factual rationalization of independent stations



Price changes are expressed as deviations from the statu-quo average price level. The order of exit corresponds to the inverse marginal cost ordering.

of the multi-address model is that heterogeneity in consumers' commuting behavior translates into heterogeneity in the distance of consumers to stations. For instance, a long commuter is more likely to drive by a large number of stations than a consumer with a short commute; a distinction that is not present in the home-location model since all consumers tend to prefer buying gasoline in their home neighborhood. This increases the ability of consumers to substitute easily toward stores that are not physically close to each other, as indicated in the cross-elasiticity regressions in Table 8. Table 9 also reveals that this feature of the commuting model dramatically reduces the ability of firms to price discriminate across consumers, as the level of markup dispersion is systematically higher under the home-location assumption. This is particularly striking in the more competitive market structure (i.e. store-level pricing in column 1), in which markups are predicted to be 30% less dispersed by the home-location model compare to the commuting model.

In order to further analyze the market power of major retail chains in this market, I compute a counter-factual rationalization of all independent chains with fewer than 10 stores. Depending on the period, this corresponds to the exit of around 25% of all active stations. For every period in which the price floor regulation is not binding, I simulate a stepwise rationalization of the industry which removes independent retailers starting with the least efficient and ending with the most efficient (defined in terms of marginal cost). After each independent is removed, the equilibrium prices are re-computed. The goal of this exercise is to study the role of independent stations in keeping prices low, and reducing the market power of large chains. Figure 7 presents the sequences of price changes for the spring of 5 different years. Each price change is expressed relative to the base case where no stations exit. The most striking feature of these graphs is that the price changes all exhibit a U-shaped pattern in which the average market price initially falls after the exit of the first independents, and then starts raising back toward or above its initial level. This shape is comes from a trade-off between market power forces that push the prices up after the exit of independent stations, and efficiency reasons that lower the equilibrium prices after the exit of high-cost and/or low valuation stores. The fact that the average price starts increasing only after the exit of the 40th or 50th station, reveals that a large number of independent retailers have high marginal costs and offer a low value product. This also implies that major chains are more efficient, tend to invest in higher value amenities, and have better locations than independent stations. Since these investments are associated with higher fixed costs, they are not accounted for in the marginal cost of stations.

7 Conclusion

In this paper I develop, and estimate a novel model of demand for spatially differentiated products, applied to retail gasoline. My approach contributes to the literature on spatial differentiation by formally modeling commuting paths as the "locations" of consumers. This extension of the standard home-address model generates substitution patterns which depends in an intuitive way on the structure of the road network and the direction of traffic flows.

The methodology combines computing tools from the transportation Geographic Information System literature, and econometric methods developed to estimate discrete choice models of demand. The model is estimated using a unique panel dataset, which covers an important period in the evolution of the gasoline retail industry. This period is characterized by a large North-American re-organization of retail networks induced by several technological innovations. This feature of the data enables us to identify the structural parameters of the model, even after controlling for important unobserved characteristics of store locations.

The results validate the modeling choices in many ways. In particular, the distribution of gasoline sales within the market is shown to be poorly correlated with the distribution of population, and significantly more so with the distribution of commuters. In fact, this correlation is negative for some specifications studied, suggesting that a large fraction of gasoline demand is generated in areas far from where people are living. This directly translates into a negative estimate of the transportation cost parameter in the traditional model, which is inconsistent with any model of spatial differentiation. Moreover, even holding the parameters fixed, the estimates indicate that the substitution patterns predicted by the commuting model are significantly different from the ones generated by a single home-location model. Since the relative substitution of locations feeds directly into predictions of mark-ups and prices, the results present a more reliable evaluation of market power in the market than the standard model.

The market equilibrium simulations performed with the estimates offer important insights for understanding the sources of market power of retail chains. First, the ability of major chains to exercise market power is tightly linked with their ability to organize their network of stores in ways which maximizes their profits given the driving behavior of consumers. In addition, the five major chains are shown to be more efficient than independent retailers. This difference is reflected by a higher quality of their stores, and lower marginal costs. These differences imply that independent retailers are weak competitors, which reduces their ability to limit the market power of major chains.

The results of the policy experiments suggest that a reduction in the number of independents could be beneficial for consumers, by lowering the average market prices. This has important implications for evaluating the usefulness of protecting independent retailers through price floors or contract restrictions. If those policies reduce the incentive of inefficient firms to exit the market, they will keep the level of prices artificially high.

Finally, the methodology developed in this paper could easily be applied in other industries. In fact, in most retail markets the single-address model is too restrictive to describe accurately the shopping decision of consumers. For instance, in grocery retail markets consumers can easily choose a store on their way to work, and therefore the model presented here could easily be applied.

In the retail banking market, consumers' work and shopping locations are complementary to their residency location (rather than substitute like here) in defining the valuation of a bank's branch network. It is thus crucial to take this multi-dimensionality into account to predict bank choices.

A Calculation of the empirical distribution of consumers

The two following subsections describe the methods used to compute the demographic statistics at the residential location level for every periods, and the distribution of individuals across origin-destination pairs.

A.1 Distribution of population across location

The main issue here is to predict the distribution of population between census years, for each census dissemination area (DA). The DAs are the smallest statistical area for which detailed demographic statistics are available. For the Quebec metropolitan area, the average population of each DA is around 500. The DAs were created by Statistic Canada for the 2001 census. In order to compute demographic statistics for previous years, we will use the fact that DAs are geographically nested in the definition of Census Tracts (CT). In particular, we will use a definition of census tract which is common to all three censuses (i.e. 1991, 1996 and 2001).

Let X_{it}^a be a variable measured at the level of aggregation $a = \{DA, CT, CMA\}$, for zone i in period t. Two types of weights are used to predict the level of X at the DA level for every periods. First, the distribution of population across DAs for the census year 2001 is obtained directly from the census aggregate tables:

$$w_{iT}^{DA}(X) = \frac{X_{iT}^{DA}}{\sum_j X_{jT}^{DA}} \tag{12}$$

The change in this weight across periods is obtained from the observed average changes at the CT level. In particular the weight of DA i for periods $t < T$ is given by:

$$w_{it}^{DA}(X) = \frac{X_{ct(i)t}^{CT}}{X_{ct(i)T}^{CT}} w_{iT}^{DA}(X) \tag{13}$$

where $ct(i)$ is a function reporting the census tract name of DA i. Assuming that the relevant population distribution within each CT is stable over time, the weight w_{it}^{DA} is an accurate representation of the relative changes in X between year t and $T = 2001$.

In order to get monthly estimates of X, I use the monthly Canadian Labour Force survey. This survey reports estimates of the adult population and the number of workers for the main Census Metropolitan Areas on a monthly basis. Rescaling the weights defined in equation 13 so that they sum to one, the predicted value for X_{it}^{DA} is obtained by:

$$X_{it}^{DA} = \frac{w_{it}^{DA}(X)}{\sum_{j \in cma(i)} w_{jt}^{DA}(X)} X_{cma(i)t}^{CMA} \tag{14}$$

where $cma(i)$ is the CMA indicator of region i, and $X^{CMA}_{cma(i)t}$ is the value of X obtained from the LF survey in period t. The previous calculation is repeated for the three key variables used in the empirical analysis: the population older than 15 years, the number of full time students, and the number of workers (number of full-time and part-time workers who are not full-time students).

A.2 Distribution of commuting trips

In order to compute the number of commuters for each pair of origin and destination zones, I used the aggregate OD matrices from the 2001 Origin-Destination survey performed by the Québec Ministry of Transportation for the Québec city CMA. The sample of individuals surveyed in the fall of 2001 correspond to $27,839$ households or $68,121$. Each individual surveyed was asked questions related to mode of transportation used and destination for four trip purposes: work, leisure, study, and shopping. The micro data generated on each trip surveyed were then aggregated using the 2001 census weights, to generate the predicted number of trips between each pair of traffic area zones (TAZ). In the 2001 survey, each OD matrix included 67 TAZs. The definition of each TAZ represents the agglomeration of one or more census tract.

To predict the traffic between each pair of DA locations, I will use two OD matrices: the OD matrix for work trips, and the OD matrix for study trips. Let ω^t_{ij} be the proportion of trips originating from TAZ i going to TAZ j, for purpose $t \in \{\text{work,study}\}$. Since each TAZ includes multiple DAs, I have to assume a distribution of trips within each TAZ. The distribution trips originating from each zone is assumed to be homogeneous across DAs within the same TAZ. This is justified by the lack of additional information, and by the fact that census boundaries are defined such that population within each CTs is as homogeneous as possible. The distribution of destinations zones within each TAZ is, on the other hand, assumed to be proportional to the distribution of employees and schools (Colleges and Universities) respectively. The distribution of employees by DAs is available only for year 2001 from the Canadian Business Summary database compiled by PCensus, while the distribution of schools by DAs is calculated using the DMTI Enhanced Points of Interest database[18]. Combining this information with the aggregate OD probabilities, we can

[18] DMTI Spatial. "Enhanced Points of Interest", version 3.1 [Electronic resource]. Markham, Ontario: DMTI Spatial, 2004.

compute the number of commuters T_{ij} for the DA pair (i, j) using the following formula:

$$T_{ij} = \sum_{p=\{\text{work,study}\}} \omega_{taz(i),taz(j)} \frac{Y_j^p}{\sum_{j' \in taz(j)} Y_j^p} X_{it}^p, \tag{15}$$

where X_{it}^p is the relevant population measure (i.e. workers or full-time students), $taz(i)$ is a function indicating the TAZ name of DA i, and Y_j^p is the number of employees in location j if the trip purpose is work, or the number schools if the trip purpose is study. Note that the previous representation implicitly assume that the geographic distribution of trips is stationary over the sample periods.

Finally, the resultant measures of traffic are aggregated into larger location areas to reduce the computation cost of the model. In particular, I aggregated to the CT level each DA for which either the size (in square kilometer) or the corresponding CT size is smaller than the median DA or CT size.

B Description of the Shortest-Path Algorithm

The set of optimal routes between each pair of origin and destination zones is computed using a version of the Dijkstra's Shortest Path Tree algorithm (see Shekhar and Chawla (2003) for an enlightening introduction to this class of algorithm). The road network is represented by a directed graph $G = (N, A)$. Where N is the set of nodes (or intersections), and A is the set of arcs (or street segments). Each segment a is a pair of connected nodes (i, j), ordered according to the direction of the arc. The time cost of traveling along each arc is given by $C = \{c_{ij} | (i, j) \in A\}$. The shortest path algorithm constructs, for every origin nodes s, a shortest path tree (SPT) $\mathcal{P}_s$ which stores the shortest path from s to every other nodes in the network. The procedure is an iterative algorithm which iterates on the cost $t(r, v)$ of traveling from s to any node v until convergences.

At any point during the iteration process, the algorithm keeps track of a list of nodes left to be examined ($frontierSet$), a list of nodes already explored ($exploredSet$), and a function $p_s(v)$ which indicates the parent node in the shortest path from s to v. At each iteration the algorithm removes the lowest cost node from the frontier set, and visit every nodes that are adjacent to this node (i.e. $adjSet(u)$). If the cost of visiting one of these nodes $w \in adjSet(u)$ is lower than the current estimate, the algorithm updates the cost function $t(s, w)$ and the path $p_s(w)$. The valid nodes are then added to the frontier set. The algorithm stops when all nodes in the network have been visited. The pseudo-code below describes the main steps of the SPT calculation.

Algorithm 1. *Shortest path three rooted at node s, on network $G(N, A)$:*

Initialization step:

$$t(s,v) = \begin{cases} \infty & \text{if } v \neq s \\ 0 & \text{otherwise} \end{cases} \qquad frontierSet^0 = \{s\} \qquad exploredSet^0 = \{\emptyset\}$$

Iteration k:

$u = \arg\min_{w \in frontierSet^k} \quad t(s,w)$

$frontierSet^k = frontierSet^{k-1} \backslash \{u\}$

$exploredSet^k = exploredSet^{k-1} \cup \{u\}$

`foreach` $w \in adjList(u)$

`if` $t(s,w) > t(s,u) + c(u,w)$ `then`

{

$t(s,w) = t(s,u) + c(u,w)$

$p_s(w) = u$

`if` $w \ni frontierSet^k \cup exploredSet^k$ `then`

$frontierSet^k = frontierSet^k \cup \{w\}$

}

`if` $frontierSet^k = \emptyset$ `then`

`stop`

`else` $k = k + 1$

The set of routes corresponding to the shortest path tree $\mathcal{P}_s$ are constructed recursively using the function $p_s(v)$. For instance the path $r(s,d) \in \mathcal{P}_s$ is an array of $n_r + 1$ nodes such that the last element is $r_{n_r} = d$, the second-last element is $r_{n_r - 1} = p_s(d)$, the k^{th}-last element is $r_{n_r - k + 1} = p_s(r_{n_r - k})$, and the first element is $r_0 = s$.

C Computation of the GMM estimator

In this section I describe the details involved in the estimation and statistical inference of the parameter vector θ, and various functions of those parameters (e.g. elasticities, willingness to travel, markups, etc). In particular, I discuss three elements of the procedure which are specific to my problem: (i) the inversion procedure, (ii) the MCMC estimation algorithm, (iii) the construction of the weighting matrix.

C.1 Inversion algorithm

In order to evaluate the objective function at a given parameter vector θ, it is necessary to invert the following system of non-linear equations:

$$\delta(\theta)_{jt} \rightarrow \ln Q_{jt}(\delta_t|\theta) = \ln q_{jt} \tag{16}$$

where $Q_{jt}(\delta_t|\theta)$ is the model predicted market share at store j in period t, and q_{jt} is the observed share. The complexity of the inversion procedure depends in general on the number of consumer types (e.g. the number of simulated consumers or the number of pairs (s,d)) and on the number of products available. To reduce the complexity of the problem, I compute $Q_{jt}(\delta_t|\theta)$ using a monte-carlo simulation approximation rather than aggregating the choice probabilities of about $36,000$ types of consumers. The simulation is conducted by drawing $S = 5,000$ consumers from the empirical distribution T^t_{sd}.

In addition, I use a Broyden's root-finding algorithm to evaluate equation 16 (see Miranda and Fackler (2002) for more details). This procedure is proven to converge significantly faster than the standard contraction mapping algorithm proposed by Berry et al. (1995). Letting $f(\delta^k) = \ln Q_{jt}(\delta_t|\theta) - \ln q_{jt}$ and I_{J_t} denotes the identity matrix of dimension J_t, the algorithm takes the following steps to find $\{\delta(\theta)_{jt}\}_{j=1,\dots,J_t}$:

1. Set the starting value for the pseudo-jacobian matrix $B^0 = I_{J_t}$ and δ^0_{jt}.

2. For iteration $k \geq 1$:

   (a) Update the vector of mean qualities:

   $$\delta^k_{jt} = \delta^{k-1}_{jt} - B^{k-1} f(\delta^{k-1}_t)$$

   (b) Update the pseudo-jacobian matrix:

   $$B^k = \begin{cases} B^{k-1} + (s-u)s' B^{k-1} * (s'u)^{-1} & \text{if } ||f(\delta^k_t)|| > ||f(\delta^{k-1}_t)|| \\ I_{J_t} & \text{Otherwise.} \end{cases}$$

   where $s = -B^{k-1} f(\delta^{k-1}_t)$ and $u = B^{k-1}\big[f(\delta^k_t) - f(\delta^{k-1}_t)\big]$.

3. Stop if $||f(\delta_t^k)|| \leq \epsilon$, repeat step 2 otherwise.

Note that contrary to standard quasi-newton algorithms, this procedure is guaranteed to convergence as $k \to \infty$. To see this, note that by fixing $B = I_{J_t}$ the procedure is equivalent to the contraction mapping algorithm of Berry et al. (1995). Step 2b uses this property of the problem by reverting to the contraction mapping algorithm each time the algorithm tends to diverge. More importantly, the algorithm typically converges in less than 10 iterations, compare to more than 100 for the contraction mapping algorithm. In fact, the algorithm typically requires almost as few iterations to converge as the Newton algorithm without requiring any matrix inversion, nor evaluating the Jacobian matrix.

C.2 MCMC algorithm

The suggestion of Chernozhukov and Hong (2003) is to rewrite the GMM objective as a pseudo-likelihood problem, and compute the posterior distribution of the parameters of interest using Monte-Carlo Markov-Chain (MCMC) methods. The key result in this paper is to show that the moments of any functions $f(\cdot)$ of the parameters, evaluated using the posterior distribution of θ, corresponds to the asymptotic distribution of $f(\hat{\theta})$, where $\hat{\theta}$ solves the GMM problem.

In order to compute empirically the posterior distribution of θ, I use the following Metropolis-Hastigs algorithm:

1. Choose θ^0 (e.g. value of θ after S simplex iterations) and compute $V_n(\theta^0)$

2. Draw $\theta^* \sim q(\theta^*|\theta^j)$ and evaluate $Q(\theta^*)$

3. Update θ^{j+1} using:

$$\begin{aligned} \theta^{j+1} &= \theta^* \quad \text{with prob. } \rho(\theta^j, \theta^*) \ , \\ &= \theta^j \quad \text{with prob. } 1 - \rho(\theta^j, \theta^*) \ . \end{aligned}$$

Where,

$$\rho(x, y) = \min\left(\frac{e^{-Q(y)}\pi(y)q(x|y)}{e^{-Q(x)}\pi(x)q(y|x)}, 1 \right)$$

In the current application I use an iid normal distribution for candidate distribution $q(x|y)$, and uninformative priors (i.e. $\pi(x) = 1$).

In order to conduct inference about the parameters, one can simply compute the mean, standard-deviations or quantiles of the empirical distribution of parameters generated from the MCMC chain. Similarly, for all or a subset of the parameters in the chain, one can compute various composite parameters of θ and evaluate their moments. For instance, if $\mathcal{P}(\theta|X)$ represents the set of parameters in the chain, we can calculate the mean of store j price elasticity as:

$$\bar{e}_{jt} = \frac{1}{M}\sum_i \frac{\partial Q_{jt}(\theta_i)}{\partial p_{jt}}\frac{p_{jt}}{Q_{jt}}$$

where $\theta_i \in \mathcal{P}(\theta|X)$ and M is the number of MCMC draws.

C.3 Construction of the weighting matrix

The weighting matrix is block-diagonal as in Petrin (2002), since the two moments are calculated from different samples.[19] In the second stage of the GMM optimization routine, the weighting matrix for the second set of moment conditions is computed using a heteroskedastic-consistent variance-covariance matrix. For the first set of moments, I compute a weighting matrix which is consistent with both with spatial and time correlations in the empirical moments. In particular following Conley (1999), the variance-covariance matrix of the first set of empirical moments V_n^1 is estimated by weighing observations according to their distance in space and time:

$$\hat{V}_n^1 = \frac{1}{n}\sum_{j,t}\sum_{k,s} K\big((jt),(ks)\big) g_{jt}^1(\hat{\theta}^1) g_{ks}^1(\hat{\theta}^1)' \tag{17}$$

where $\hat{\theta}^1$ is a consistent estimate of the parameters obtained using $V_n^1 = \tilde{W}^{1T}\tilde{W}^1$, $K\big((jt),(ks)\big)$ is a distance kernel density determined by three distance metric: (i) the time difference between t and s, (ii) the physical distance between j and k, and (iii) the common street indicator variable $S(j,k)$. To combine these three metrics, I use a distance kernel which is the product of two normal densities if two stores share a street and the time lag between two periods is no longer than 6 periods (i.e. 1 year). More specifically $K\big((jt),(ks)\big)$ is given by:

$$K\big((jt),(ks)\big) = \begin{cases} \phi\big((t-s)/\sigma_T\big) \times \phi\big(d(j,k)/\sigma_d\big) \times S(j,k) & \text{if } t-s \leq 6 \text{ and } t \geq s \ , \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

where σ_T and σ_d are the standard deviations of the time difference and driving distance variables, and $\phi(\cdot)$ is the standard normal density.

[19] See Imbens and Lancaster (1994) for further details on the estimation of micro-econometric models with macro moment conditions.

References

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2005, May). Econometric tools for analysing market outcomes. mimeo, Harvard University.

Anderson, S., A. de Palma, and J.-F. Thisse (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press.

Berkowitz, M., N. T. Gallini, E. J. Miller, and R. A. Wolfe (1990, May). Disaggregate analysis of the demand for gasoline. *The Canadian Journal of Economics 23*(2), 253–275.

Berry, S. (1994). Estimating discrete choice models of product differentiation. *Rand Journal of Economics 25*, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society 63*(4), 841–890.

Berry, S., J. Levinsohn, and A. Pakes (1999). Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review 89*(3), 400–430.

Brambilla, I. (2004, October). A customs union with multinational firms: The automobile market in argentina and brazil. Yale University.

Bresnahan, T. F. (1987). Competition and collusion in the american automobile industry: The 1955 price war. *The Journal of Industrial Economics 35*(4, The Empirical Renaissance in Industrial Economics), 457–482.

Chernozhukov, V. and H. Hong (2003). An mcmc approach to classical estimation. *Journal of Econometrics 115*, 293 346.

Claycombe, R. (1991). Spatial retail markets. *International Journal of Industrial Organization 9*, 303–313.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics 92*, 1–45.

Davis, P. (2006, Winter). Spatial competition in retail markets: Movie theaters. *Rand Journal of Economics*.

DMTI-Spatial (2004). *CanMap RouteLogistics, version 8.2 [Electronic resource]*. Markham, Ontario.

Dubé, J.-P. (2005). Product differentiation and mergers in the carbonated soft drink industry. *Journal of Economics and Management Strategy 14*(4), 879–904.

Dubin, J. A. and D. L. McFadden (1984, March). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica 52*(2), 345–362.

Eckert, A. and D. S. West (2005). Rationalization of the retail gasoline station networks in canada. *Review of Industrial Organization 26*, 1–25.

Hastings, J. (2004). Vertical relationships and competition in retail gasoline markets: Empirical evidence from contract changes in southern california. *American Economic Review*.

Hastings, J. and R. Gilbert (2002, June). Market power, vertical integration, and the wholesale price of gasoline. University of California Energy Institute Power Working Paper, No. PWP-084.

Imbens, G. and T. Lancaster (1994, October). Combining micro and macro data in microeconometric models. *The Review of Economic Studies 61*(4), 655–680.

Lach, S. (2002). Existence and persistence of price dispersion: An empirical analysis. *The Review of Economics and Statistics*.

Manez, J. and M. Waterson (2001). Multiproduct firms and product differentiation: a survey. working paper, University of Warwick.

Manuszak, M. D. (2001). The impact of upstream mergers on retail gasoline markets. Working Paper, Carnegie Mellon University.

McManus, B. (2006). Nonlinear pricing in an oligopoly market: The case of specialty coffee. Forthcoming, Rand Journal of Economics.

Miranda, M. J. and P. L. Fackler (2002). *Applied computational economics and finance*. MIT Press.

MTQ (2002). La mobilité des personnes dans l'agglomération de québec. somaire des résultats de l'enquète origine-destination 2001. Technical report, Ministère des Transports du Québec et Réseau de Transport de la Capitale.

Nevo, A. (2000). Mergers with differentiated products: The case of the ready-to-eat ceral industry. *The Rand Journal of Economics 31*(3), 395–421.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica 69*(2), 307.

Noel, M. (2005). Edgeworth price cycles: Evidence from the toronto retail gasoline market. University of California at San Diego.

Oppenheim, N. (1995). *Urban Travel Demand Modelling*. Wiley-Interscience.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy 110*, 705.

Petro-Canada (1996, September). Government regulation and intervention in the downstream oil industry. Technical report, Petro-Canada. Submitted to the Government of the Province of Québec.

Pinkse, J. and M. Slade (2004). Mergers, brand competition and the price of a pint. *European Economic Review 48*, 617–643.

Pinkse, J., M. E. Slade, and C. Brett (2002). Spatial price competition: A semiparametric approach. *Econometrica 70*(3), 1111–1153.

Raith, M. (1996). Spatial retail markets commuting consumers. *International Journal of Industrial Organization 14*, 447–463.

Shekhar, S. and S. Chawla (2003). *Spatial Databases: A Tour*.

Smith, H. (2004). Supermarket choice and supermarket competition in market equilibrium. *The Review of Economic Studies 71*(1), 235–263.

Theriault, Marius, V. M.-H. L.-G. M. and D. Leroux (1999). Modelling commuter trip length and duration within gis: Application to an o-d survey. *Journal of Geography Information and Decision Analysis 3*(1), 40–56.

Thomadsen, R. (2004). The effect of ownership structure on prices in geographically differentiated industries. *Rand Journal of Economics (forthcoming)*.

Yatchew, A. and A. No (2001). Household gasoline demand in canada. *Econometrica 69*, 1697–1708.