March of the Chains: Herding in Restaurant Locations*

Nathan Yang†
Yale School of Management

October 22, 2012

Abstract

Does learning from others generate retail clusters? Uncertainty about new markets provides an opportunity for learning from others, where incumbents' past stay/exit decisions are informative to potential entrants. The setting is Canada's fast food industry from 1970 to 2005, where I present a new estimable dynamic oligopoly model of entry/exit with unobserved heterogeneity, common uncertainty about profitability, learning through entry, and learning from others. With the estimated model, I find that learning induces retailers to herd into markets that others have previously done well in, avoid entering markets that others have previously failed in, and for some, strategically delay entry. Finally, I show that entry deterrence may come at a cost, in the form of added risk from entering early.

Keywords: Agglomeration, dynamic discrete choice game, market structure, retail industry.

*I am thankful for supervision from Victor Aguirregabiria, and careful guidance from my thesis advisors. I gratefully acknowledge funding from Social Sciences and Humanities Research Council Canada and the NET Institute (www.netinst.org). The data on Canada's fast food industry were collected with the aide of patient staff from the Toronto Reference Library. This paper is from the second chapter of my thesis done while at the University of Toronto.

†Email: nathan.yang@yale.edu.

1 Introduction

Retail managers are often faced with the difficult decision of where to place their stores.[1] Such decisions are challenging because of the uncertainty retailers face; especially so if this uncertainty cannot be fully resolved via market research. For instance, American retailers may be uncertain about a market's tastes (Bell and Shelman, 2011), anti-American sentiment (Beamish, Jung, and Kim, 2011), and health consciousness (Lawrence, Requejo, and Graham, 2011). In some cases, it is only by diving into a market that such uncertainty would be resolved (i.e., *learning through entry*). But upon entering a market, subsequent stay/exit decisions are publicly seen, and thus, prospective entrants can infer market profitability based on such observations (i.e., *learning from others*). In fact, it has been conjectured by Toivanen and Waterson (2005) in their study of Burger King and McDonald's in the United Kingdom, as well as Shen and Xiao (2012) in their study of Kentucky Fried Chicken and McDonald's in China, that learning from others may explain the commonly observed clustering of seemingly rival retail chains.[2] Similar patterns have also been documented for the retail banking industry (Damar, 2009; Feinberg, 2008), as well as department stores (Vitorino, 2008).

In past literature about retail, researchers have posited unobserved heterogeneity and demand externalities as typical explanations for retail clustering. A nearby mall, local attraction, or highway exit can easily generate retail agglomeration among rivals (Orhun, 2012; Thomadsen, 2007), as can restrictive retail zoning provisions (Datta and Sudhir, 2011) - both factors pointing towards unobserved heterogeneity. Alternatively, a store may generate demand externalities for neighboring rivals if its presence helps draw in additional consumer traffic (Datta and Sudhir, 2011; Eppli and Benjamin, 1994; Konishi, 2005), or if its close proximity can credibly soften price competition via market segmentation or cannibalization[3] concerns (Thomadsen, 2010; Zhu, Singh and Dukes, 2011).

Despite the well-developed theoretical literature on social learning and learning-from-others,[4] empirical research on retail agglomeration has largely overlooked the idea that if managers face uncertainty about market profitability, then they may have an incentive to take advantage of any

[1] In general, a key part of retail marketing is identifying demand across different regions (Kotler and Keller, 2004).

[2] Other settings where information externalities are possible include: computer purchase decisions (Goolsbee and Klenow, 2002), farming technology adoption (Conley and Udry, 2010), macroeconomic policy choice (Buera, Monge-Naranjo, and Primiceri, 2010), kidney adoption (Zhang, 2010), momentum effects in sequential elections (Knight and Schiff, 2007), movie sales (Moretti, 2010), SARS risks (Bennett, Chiang, and Malani, 2011), Twitter adoption among politicians (Chi and Yang, 2010), and word of mouth in online book sales (Chevalier and Mayzlin, 2006).

[3] For example, if a market has one McDonald's outlet, and one Burger King outlet, the entry of one additional McDonald's outlet can actually benefit both chains. The additional McDonald's outlet will induce McDonald's to price less competitively so as to avoid cannibalizing its original store's sales.

[4] Literature that builds on Caplin and Leahy (1998), and Chamley and Gale (1994).

information that can possibly be revealed when an existing and informed chain decides to stay or exit a market. My objective is to understand how these externalities will affect an industry, and whether they contribute to behavior consistent with clustering. The setting for my analysis is Canada's fast food industry, where I study the entry/exit decisions of the five major fast food chains in Canada - A & W, Burger King, McDonald's, and Wendy's, along with the Canadian chain Harvey's - from the industry's beginning[5] around 1970 to 2005 - across small geographic markets nested within all Canadian cities (Section 2).

Section 3 presents a descriptive empirical regularity that shares similarities with previous studies (Shen and Xiao, 2012; Toivanen and Waterson, 2005). In particular, I find that the incumbency status of a chain has a positive effect on its rivals' decisions to enter a local market, even when (time-varying) unobserved heterogeneity is accounted for. A consistent theme throughout this empirical analysis is that fast food chains tend to follow their rivals into markets. These patterns are certainly suggestive of clustering. Not surprisingly, the fast food industry has become an increasingly popular laboratory for studying retail agglomeration (Thomadsen, 2007, 2010; Toivanen and Waterson, 2005).

This paper introduces a new estimable dynamic oligopoly model of entry[6] that allows for common uncertainty, permanent unobserved heterogeneity, and information externalities in Section 4. The basic idea of the model is that *ex ante*, chains face uncertainty about market profitability. This uncertainty is resolved after entry, as incumbents will become informed by observing the true market profitability via realized revenues. The decisions of incumbents will be made without uncertainty about market profitability, thereby giving rivals who have not yet entered an opportunity to learn from these observed stay/exit decisions. The model I introduce is unique in its interpretation of learning in that retailers face some uncertainty that can only be resolved either by their own entry, or through inference from the observable past decisions of rivals. Learning behavior has long been embedded in models of industry dynamics, dating back to Jovanovic (1982). The main interpretation of uncertainty in these models is regarding a firm or its rival's cost advantage. Recent empirical applications of such models include the study of predatory behavior by incumbent airlines (Kim, 2009), and competition in China's microwave industry (Shen and Liu, 2012).

The inclusion of dynamics is important for this empirical study of observational learning. By

[5] Other studies that investigate empirical patterns in retail industry dynamics are Eckert and West (2008), Kosová and Lafontaine (2010), and Shen (2010) to name a few. Many of these studies are motivated by the theoretical framework of Jovanovic (1982).

[6] Other papers that estimate dynamic/static structural models pertaining to retail chains include Ellickson, Houghton and Timmins (2010), Holmes (2011), Jia (2008), Nishida (2008), Orhun (2012), Suzuki (2010), Toivanen and Waterson (2005, 2011), Varela (2010), and Zhu, Singh and Duke (2011).

allowing the retailers in my model to be forward looking, they can react appropriately to information externalities. For instance, a potential entrant may have an incentive to strategically delay entry as dictated by the option value of waiting (Chamley, 2004), while an incumbent may anticipate increased competition in the future as its decision to stay in the market may cause rational herding. Under a static setting, such behavior is restricted.

Identification of the model's parameters is discussed in Section 5, where key issues pertain to how learning can be separately identified from unobserved heterogeneity and strategic interactions. The intuition behind identification of learning is as follows: unlike unobserved heterogeneity and strategic interactions, a retailer will react differently to its rival's past decision to stay/exit depending on whether the retailer is an uninformed potential entrant, or an informed incumbent; therefore, learning, unobserved heterogeneity, and demand spillovers will come into play for an uninformed retailer, while only unobserved heterogeneity and demand spillovers will be relevant for an informed retailer. In fact, a simplified version of the model allows me to derive formally a novel differences-in-differences (DID) approach that can be used as a preliminary test for learning.

I later describe how the new dynamic model of entry/exit is estimated using a combination of grid search, Nested Pseudo Likelihood (Aguirregabiria and Mira, 2007), and Expectation-Maximization (Arcidiacono and Miller, 2011). The structural estimates presented in Section 6 demonstrate that the fast food chains do indeed face uncertainty, which is a necessary condition for learning. Subsequent counterfactual analysis allows me to assess whether or not uncertainty (and thus learning) can in fact induce retailers to cluster. By simulating a counterfactual equilibrium for which the degree of uncertainty is set to zero, and comparing this equilibrium with the actual one, I find that an industry is *more agglomerated* when uncertainty is present, as retailers are more inclined to *follow* successful rival incumbents into the same markets, and to *avoid* entering markets in which rival incumbents failed. Furthermore, I find that certain retailers have an incentive to strategically delay their entry, as predicted by theoretical models of social learning. Finally, I show that firms that successfully preempt markets from competitors face a more pronounced trade-off with added market risk when there is uncertainty and learning from others.

2 Data and industry

2.1 Canada's hamburger fast food industry

This study investigates local competition among fast food outlets that primarily serve hamburgers. I focus my attention on the five largest chains operating in Canada: A & W, Burger King, Harvey's, McDonald's and Wendy's. In Canada, no other chains with national presence entered the industry but failed as a whole. Hence, the set of five chains I look at is very representative of hamburger fast food chains in Canada. Note that there exist quick-service outlets that do not serve hamburgers, such as Kentucky Fried Chicken, Subway, and Taco Bell, which I leave out from my analysis largely because the products offered by hamburger chains are likely to be more substitutable with one another. Furthermore, these chains are late entrants into Canada relative to the hamburger chains. Although Kentucky Fried Chicken was available as early as 1953, it was primarily served through convenience stores until the 1980s. Subway's first outlet in Canada was opened in 1986, while Taco Bell's first outlet in Canada was opened in 1981.

Since 1970, Canada has become a very important foreign market for American retail chains. Canada provides American chains a real growth option,[7] without the risk associated with more exotic markets overseas (Holmes, 2010). Not surprisingly, American chains tend to launch in Canada first before they expand to other countries (Smith, 2006); this strategy is a general phenomenon seen in the entire retail industry. In fact, McDonald's was largely motivated to expand globally after its success in Canada (Love, 1995). Using Canada as a stepping stone, all four of the American chains are currently active players in the global fast food industry. Today, McDonald's has almost 31,000 outlets around the world, Burger King has 4,000 outlets, then A & W follows with about 700, and 400 for Wendy's internationally. The largest domestic chain, Harvey's, boasts a store count of over 200 outlets in Canada.

Many of these franchises were founded in the United States prior to 1970. A & W in 1956, Burger King in 1952, McDonald's in 1952, and Wendy's in 1969; Canada's chain Harvey's was founded in 1959. The first American chains to set up in Canada were A & W (1956), and McDonald's (1967). Although their relative standings have changed over time, these five chains are still the most dominant forces in Canada's fast food industry today.

2.2 Local market definition and observable characteristics

I consider a Forward Sortation Area (FSA) as a local market. FSA designations are defined as the first three digits of a postal code and are loosely based on population. They are on average 1.8 square miles in many Canadian cities, and thus, comparable to American Census Tracts. Note that the markets I use are smaller than those used in other studies on retail competition and agglomeration. For example, Toivanen and Waterson (2005) use Local Authority Districts in the

[7] Franchised chain growth in Canada is still markedly smaller than growth in America. Kosová and Lafontaine (2010) show that growth is about 29 percentage points lower in Canada as compared to the States.

Table 1: Coverage of CMAs in sample.

Province	Cities
Alberta	Calgary, Edmonton
British Columbia	Vancouver, Victoria, Kelowna, Abbotsford
Manitoba	Winnipeg
New Brunswick	Moncton, Saint John
Newfoundland	St. John's
Nova Scotia	Halifax
Ontario	Toronto, Ottawa, Hamilton, London, Windsor, Niagra Falls, Peterborough, Guelph, Kitchener, Kingston, Oshawa, Barrie, Brantford, Sudbury, Thunder Bay
Saskatchewan	Saskatoon, Regina

United Kingdom, which are equivalent to cities. Ellison, Glaeser, and Kerr (2010) use Primary Metropolitan Statistical Areas, Counties, and States; all of which are larger than FSAs. Finally, Shen and Xiao (2011) focus on city markets in China. I find 608 FSA markets based on the cities used in my sample. Because this study is focused on understanding retail clustering, we need a market definition that is as small as possible. One nice feature of the FSA market definition is that they were established well before the fast food chains entered Canada, and that all of the FSA market definitions in my sample have not undergone changes over time.

The FSA regions I sample are those nested within Canada's Census Metropolitan Areas (CMAs), or loosely speaking, all cities in Canada. As shown in Table 1, my set of CMAs covers all of the provinces in Canada, although a large proportion of them are concentrated in the province of Ontario. I focus on FSAs within CMAs since a majority of Canadians live in cities. For instance, British Columbia and Ontario have 85 percent of residents living in cities, Alberta has 82 percent, Quebec has 80 percent, and Manitoba has 71 percent.

I later match the market structure data with proxies for market size. The first variable is FSA population, which is available from the Census Profiles for the years 1986, 1991, 1996, 2001 and 2006. I impute the missing years using the inferred population growth rates. Table 2 summarizes the market characteristics that I use for the analysis. Additional information from the Census includes the average income (in Canadian dollars) of an FSA market, the average property value for each market, as well as the percentage of residents who work in/out of an FSA market. Property value is used as a proxy for the cost of purchasing a location to house a fast food outlet.

I supplement the Census data with the Small Area Retail Trade Estimators (SARTE). These

Table 2: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.
Population (persons)	20,333	11,206	44	89,686
Population density (persons per sq km)	2,344.034	3,487.339	0.186	144,908.844
Total sales (billion CDN)	1.087	1.100	0.001	9.155
Total retail locations	449	377	3	2,904
Income (dollars)	62,889.79	23,181.81	12,611.58	469,121
Property value (million CDN)	0.320	0.225	0.014	3.340
University (dummy)	0.054	0.225	0	1
Proportion work in same FSA	0.58	0.287	0	1
N				21,528

data contain information on annual total retail sales and total number of retail locations in a given FSA region, which should partially control for heterogeneity in retail activity across markets. All retail locations that belong to chains with at least 4 stores are included in this data. SARTE is the most reliable dataset of retail sales at such a disaggregated level. However, its time series variation might not be reliable.[8] Consequently, I use the 2002 survey and use it as a control for permanent cross-sectional heterogeneity. As a final control for market profitability, I include a dummy variables which indicates whether an FSA contains an accredited university. Given that fast food chains often target young adults in their ads, I can identify whether they actually locate near these populations. Note that all of the universities in my sample were established well before 1970.

My sample contains a number of markets which may be not be conducive to retail. For example, zoning regulation may prohibit retail from operating in certain FSAs; alternatively, certain FSAs may be very undeveloped and deserted. To rule out these markets, I exclude from my sample markets that have either zero retail sales/locations, population, or income.[9] After these exclusions, the number of observations is reduced from 21,888 to 21,528. As population and income changes over time, I only include market-time observations of years for which population and income are positive.

[8] Unlike the households surveyed in the Canadian Census, each chain establishment operating in a particular FSA is not surveyed. Instead, a sample of them are chosen; and each year, this sample is different. Furthermore, data from multiple years is hard to match as the FSAs covered in one year differs from FSAs covered in another year. Thus,I chose the year that had the best coverage.

[9] A more direct way of identifying retail markets would be to use geographic zoning data as in Datta and Sudhir (2011). Unlike the United States, high quality zoning data is hard to find for Canadian municipalities.

Table 3: Tabulation of the lagged active statuses.

Active two periods ago	0		1	
Active one period ago	0	1	0	1
A & W	16,904	264	96	3,408
Burger King	18,092	200	37	2,343
Harvey's	17,943	228	70	2,431
McDonald's	11,471	449	2	8,750
Wendy's	18,448	177	28	2,019

2.3 Entry and exit data

I turned to archived phone books at the City of Toronto's Reference Library for information about each outlet's location, time of opening, and if applicable, time of closing. There, I am able to find series of phone books, from 1970 to 2005 for virtually all 33 of the CMAs in Canada. Searches based on CMAs are necessary as the library does have complete series for the smaller Census Areas (CA's). Note that the CMAs of Sherbrooke, Saguenay and Trois-Rivieres are left out because of missing phone directories over certain time intervals. This method allows me to identify:

1. **Opening year:** The first year in which a particular outlet is listed in the phone directory.
2. **Closing year:** The last year in which a particular outlet is listed in the phone directory.
3. **Location:** The exact address of each outlet.

Outlets that first appear in the 1970 phone books may have opened in earlier years. To investigate whether this cut-off is appropriate, I look at the older phone directories (1950-1970) for some cities. With the exception of a few A & W and Harvey's outlets, very few in my sample actually opened before 1970. Each address is later geocoded and assigned a 6-digit postal code using Geocoder.ca. For each relevant FSA, I identify whether or not a chain is active in a particular FSA; a chain is defined to be active if it has at least one active store in the market.

Figure 1 highlights the amount of variation in both entry and exit over time. Furthermore, there is quite a lot of variation in the sequence of entry/exit decisions, as indicated in Table 3. In general, the fast food industry is quite dynamic.

Table 4 shows that each FSA can contain upwards of 9 outlets for a given chain. However, the fast food chains typically operate either 0 or 1 outlet in each market. Fewer than 5% of my

Figure 1: Total number of outlets opened/closed in Canada over time.

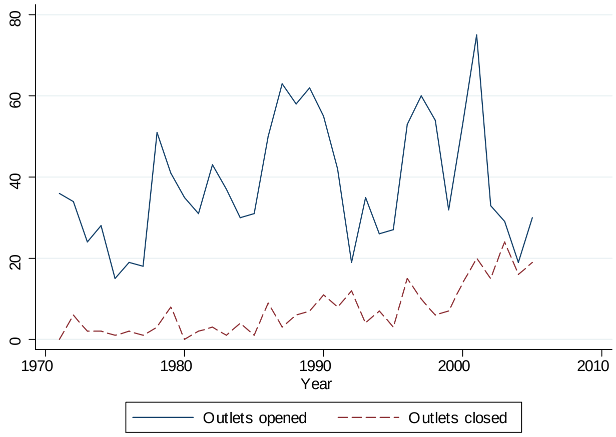


Table 4: Tabulation of market-time observations that contain 0, 1, ..., 9 outlets belonging to each of the chains.

	A & W	Burger King	Harvey's	McDonald's	Wendy's
0	18,018	19,182	19,070	12,192	19,539
1	3,126	2,505	2,536	7,027	2,174
2	508	188	228	1,891	142
3	160	13	46	536	28
4	67	0	6	142	5
5	9	0	2	55	0
6	0	0	0	28	0
7	0	0	0	9	0
8	0	0	0	5	0
9	0	0	0	3	0

market-time observations have a chain operating more than 1 outlet. Note that eventually, all FSAs contain at least one active chain by the end of my sample.

Also, the chains in general differ in terms of their entry timing (Table 5). We see that A & W and McDonald's typically enter first. Burger King, Harvey's, and Wendy's are more often than not followers into markets. In general, there is a lot of variation in terms of the timing of their entry (Figure 2). Furthermore, we get variation in the timing of exit for the retailers, as highlighted in Figure 3; the timing of exit appears to be spread out quite well, suggesting no deterministic patterns in exit due to franchisee contract renegotiations.

There are a handful of markets that were already occupied at the beginning of my sample in 1970. To see whether these markets are inherently different from markets that were occupied after

Figure 2: Histogram of entry years.

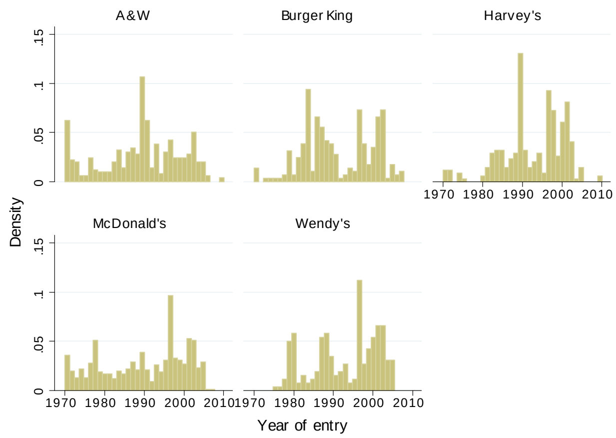


Figure 3: Histogram of the exit years.

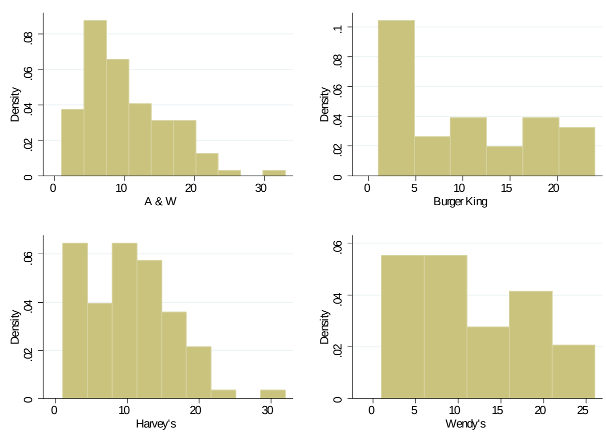


Table 5: Tabulation of the total number of markets that a chain was the (unique) first entrant.

Chain	First entrant
A & W	100
Burger King	50
Harvey's	65
McDonald's	334
Wendy's	34

1970, I calculate the mean and variance for the main variables for two sub-samples. The first sub-sample is for markets that were occupied in 1970, and the second sub-sample is for markets that were occupied after 1970. Table 6 presents the summary statistics, and in general, there are no obvious differences between these two sub-samples. It is worth noting that the markets that were first occupied in 1970 do not appear to be systematically better than markets that were explored later on.

Table 6: Summary statistics for markets that were occupied in 1970, and for markets that were occupied after 1970.

	Occupied 1970		Occupied after 1970	
Variable	Mean	Std. Dev	Mean	Std. Dev
Population (persons)	21,144	7,433	23,895	12,809
Population density (persons per sq km)	2,892.93	3,276.488	1,615.26	2,271.38
Total sales (billion CDN)	1.410	1.160	2.330	1.170
Total retail locations	483	364	850	408
Income (dollars)	57,579	14,082.81	55,518.77	18,571.69
Property value (million CDN)	0.322	0.168	0.259	0.161

3 Reduced-form evidence on clustering

The primary objective of this section is to verify a positive relationship between rivals' incumbency statuses, and one's own decision to enter/stay in a market. Evidence of such positive relationships would confirm the presence of retail clustering, and motivate further analysis to better understand the role that learning plays. A complication though is that establishing this relationship is akin to finding evidence of state dependence with unobserved heterogeneity as a confounding factor. The difference here is that not only your past state, but also your rivals' past states may matter. The econometric specification I wish to estimate for each fast food chain is thus

$$\Pr(a_{imt} = 1|\boldsymbol{a}_{mt-1}, \boldsymbol{Z}_{mt}) = \Phi(\alpha_i + \boldsymbol{Z}_{mt}\boldsymbol{\beta}_i + \sum_{j \neq i} \gamma_{ij} a_{jmt-1} + \rho_i t + \eta_m + \varsigma_i t \cdot \eta_m) \tag{1}$$

where a_{imt} is a binary choice variable that equals 1 if chain i is active in market m at time t, $\boldsymbol{Z}_{mt}$ are (time-varying) exogenous market characteristics, $\boldsymbol{a}_{mt-1} = \{a_{jmt-1}\}_j$ is the vector of past decisions, and the set of parameters $\{\gamma_{ij}\}$ captures state dependence effects. In particular, each potential spillover effect is represented by γ_{ij} for all $i \neq j$; this is the effect I am interested in estimating. Finally, the time trend is captured by ρ_i. The main complication associated with estimating this specification is the unobserved heterogeneity, captured by η_m. Because the panel data is long, I estimate the market fixed effect by including 608 market dummies into the specification. The interaction between time and the market fixed effect, $t \cdot \eta_m$, captures a restrictive form of time-varying unobserved heterogeneity. In particular, it captures time-varying unobserved heterogeneity that grows over time, such as a growth in shopping centers that draw in traffic.

Table 7 provides the first set of evidence in favor of some form of retail clustering: A & W's decision to be active is positively affected by Burger King and Wendy's incumbency status; Burger King's decision to be active is positively affected by McDonald's and Wendy's incumbency status; Harvey's decision to be active is positively affected by Burger King and McDonald's incumbency status; and Wendy's decision to be active is positively affected by A & W, Harvey's and McDonald's incumbency status. We get similar results if we use entry decisions in place of decisions to be active as highlighted in Table 8.

4 Dynamic oligopoly model of entry and exit with learning

4.1 Basic set-up

There are J chains, indexed by $i \in \{1, ..., J\}$. Time is discrete and indexed by t. Every period, the chains have to decide at the same time, whether or not to be active in a market m. Let $a_{imt} \in \{0, 1\}$ indicate whether chain i is active ($a_{imt} = 1$) or not active ($a_{imt} = 0$) during time t. Each chain's objective is to maximize the discounted payoffs $\sum_s^\infty \beta^{t+s} \Pi_{imt+s}$, where Π_{imt+s} is the one-shot payoff of firm i at period $t + s$, and $\beta \in (0, 1)$ is the discount factor. Choosing not to be active at time t yields a one-shot payoff of zero. Being active in a market yields a one-shot payoff:

$$\Pi_{imt}(a_{imt} = 1) = S_{mt}\theta_{1i} + \sum_{j \neq i} \theta_{2ij} a_{jmt} - FC_i - (1 - a_{imt-1})EC_i + \omega_m - \varepsilon_{imt}. \tag{2}$$

Here, market size is denoted by $S_{mt} = \boldsymbol{Z}_{mt}\boldsymbol{\beta}$, where $\boldsymbol{Z}_{mt}$ are observable market characteristics. The parameter θ_{1i} captures a chain specific brand effect; in other words, how effective a chain is at turning potential demand into realized sales, either through superior brand recognition or advertising campaigns. Furthermore, an active firm's variable profits depends on whether its competitors are also active in the market, as captured by θ_{2ij}. This specification for reduced form profits is similar to Vitorino (2008); furthermore, I make no assumption that $\theta_{2ij} < 0$ must hold. There are also entry and fixed costs, denoted by EC_i and FC_i respectively. Building on Seim's (2006) incomplete information framework, I assume that each chain receives a privately known and idiosyncratic

Table 7: Evidence of clustering based on the chains' decision to be active in market.

	(1) A & W	(2) Burger King	(3) Harvey's	(4) McDonald's	(5) Wendy's
A & W incumbent	3.952***	0.0712	0.0946	0.0541	0.305***
	(0.0709)	(0.0897)	(0.0894)	(0.0875)	(0.0910)
Burger King incumbent	0.363***	4.443***	0.247*	0.214	0.0169
	(0.0990)	(0.119)	(0.108)	(0.137)	(0.124)
Harvey's incumbent	0.00462	0.186	4.231***	-0.0241	0.294**
	(0.0939)	(0.102)	(0.0916)	(0.122)	(0.109)
McDonald's incumbent	0.0614	0.181*	0.364***	4.621***	0.481***
	(0.0715)	(0.0817)	(0.0745)	(0.328)	(0.0841)
Wendy's incumbent	0.385***	0.273*	0.0558	0.0851	4.617***
	(0.102)	(0.114)	(0.109)	(0.168)	(0.137)
A & W age	-0.0218***	0.0134*	-0.0155*	0.0253***	0.00338
	(0.00551)	(0.00652)	(0.00676)	(0.00695)	(0.00695)
Burger King age	-0.0130	-0.0438***	-0.00280	-0.00952	0.0263**
	(0.00907)	(0.00996)	(0.00943)	(0.0153)	(0.00991)
Harvey's age	0.0179*	0.00462	-0.0432***	0.0151	-0.00875
	(0.00817)	(0.00882)	(0.00798)	(0.0114)	(0.0105)
McDonald's age	0.00392	0.00783	0.00539	0.106	0.00252
	(0.00440)	(0.00477)	(0.00457)	(0.0832)	(0.00509)
Wendy's age	-0.00866	-0.00539	0.00933	-0.0120	-0.0501***
	(0.00934)	(0.0100)	(0.00924)	(0.0160)	(0.0109)
log(Population)	0.00598	-0.0472	0.0667	0.0815*	0.102*
	(0.0314)	(0.0374)	(0.0344)	(0.0332)	(0.0434)
log(Population density)	0.0123	0.0520*	-0.0235	0.00725	-0.0438*
	(0.0173)	(0.0220)	(0.0192)	(0.0172)	(0.0214)
log(Income)	-0.0871	-0.0368	-0.334***	-0.211*	-0.0778
	(0.0849)	(0.100)	(0.0860)	(0.0905)	(0.105)
log(Property value)	-0.149**	-0.226***	0.148**	-0.000932	-0.0899
	(0.0516)	(0.0610)	(0.0542)	(0.0531)	(0.0634)
University	0.187*	-0.0218	-0.0884	0.0322	-0.0929
	(0.0884)	(0.117)	(0.113)	(0.105)	(0.130)
Constant	0.100	0.696	-0.902	-0.652	-1.591
	(0.903)	(1.053)	(0.898)	(0.903)	(1.146)
Observations	20930	20930	20930	20930	20930
BIC	3517.7	2538.3	3116.2	3795.2	2238.2

Clustered standard errors (by FSA) in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 8: Evidence of clustering based on the chains' decision to enter a market.

	(1) A & W	(2) Burger King	(3) Harvey's	(4) McDonald's	(5) Wendy's
A & W incumbent		0.0531	0.128	0.0532	0.333***
		(0.0996)	(0.0999)	(0.0880)	(0.0983)
Burger King incumbent	0.287*		0.168	0.214	-0.0711
	(0.120)		(0.126)	(0.138)	(0.144)
Harvey's incumbent	0.0177	0.201		-0.0312	0.209
	(0.112)	(0.116)		(0.123)	(0.120)
McDonald's incumbent	0.0594	0.185*	0.398***		0.539***
	(0.0839)	(0.0879)	(0.0824)		(0.0888)
Wendy's incumbent	0.421***	0.303*	0.141	0.0800	
	(0.121)	(0.127)	(0.133)	(0.170)	
A & W age	-0.0146	0.0152*	-0.00664	0.0255***	0.00305
	(0.00862)	(0.00733)	(0.00735)	(0.00698)	(0.00767)
Burger King age	-0.0202	-0.191*	0.0155	-0.00935	0.0324**
	(0.0123)	(0.0855)	(0.0112)	(0.0154)	(0.0113)
Harvey's age	0.0205*	0.00384	-0.0315*	0.0153	0.00146
	(0.00964)	(0.0104)	(0.0160)	(0.0115)	(0.0116)
McDonald's age	-0.000225	0.00883	0.00876	-0.149	0.00439
	(0.00533)	(0.00538)	(0.00532)	(0.234)	(0.00549)
Wendy's age	-0.00278	-0.00157	-0.00692	-0.0119	-0.0629**
	(0.0121)	(0.0122)	(0.0121)	(0.0162)	(0.0237)
log(Population)	0.0185	-0.0472	0.0902*	0.0853*	0.0699
	(0.0373)	(0.0412)	(0.0398)	(0.0336)	(0.0446)
log(Population density)	0.0240	0.0640*	-0.0398	0.00800	-0.0435
	(0.0202)	(0.0251)	(0.0215)	(0.0173)	(0.0230)
log(Income)	-0.116	0.00121	-0.275**	-0.207*	-0.104
	(0.103)	(0.113)	(0.102)	(0.0912)	(0.114)
log(Property value)	-0.158**	-0.201**	0.147*	0.00926	-0.0524
	(0.0606)	(0.0678)	(0.0612)	(0.0536)	(0.0694)
University	0.176	0.0518	-0.0570	0.0274	-0.0965
	(0.106)	(0.129)	(0.124)	(0.105)	(0.143)
Constant	0.284	-0.165	-1.729	-0.864	-1.484
	(1.063)	(1.164)	(1.034)	(0.912)	(1.219)
Observations	17278	18432	18309	11819	18759
BIC	2647.1	2129.4	2422.1	3739.1	1928.7

Clustered standard errors (by FSA) in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

shock ε_{imt}, which one may interpret as some form of manager/franchisee ability. Finally, ω_m is a market-specific component that is unknown to the econometrician.

4.2 Unobserved heterogeneity

Unlike standard models of dynamic oligopoly, I allow retailers to have heterogeneous (and evolving) beliefs about ω_m. Ultimately, the posterior beliefs about ω_m will help capture the learning process. To begin, I now characterize retailers as being either informed or uninformed, conditional on their information set $\mathbf{\Omega}_{imt}$. I will describe in greater detail the elements of Ω_{imt} in the next section.

With posterior probability λ_{imt}, chain i is uninformed (i.e., faces uncertainty) in market m and time t, and with probability $1-\lambda_{imt}$, the chain is informed (i.e., does not face uncertainty). Given these beliefs, the unobserved market characteristic is defined as:

$$\omega_m = \begin{cases} \eta_m(1+\sigma_i) & \text{w.p.} \quad \lambda_{imt} \\ \\ \eta_m & \text{w.p.} \quad 1-\lambda_{imt} \end{cases} . \tag{3}$$

Therefore, retailer i's subjective belief about ω_m is

$$\begin{aligned} E_i(\omega_m|\mathbf{\Omega}_{imt}) &= \lambda_{imt}\eta_m(1+\sigma_i) + (1-\lambda_{imt})\eta_m \\ &= \eta_m + \lambda_{imt}\eta_m\sigma_i \end{aligned} \tag{4}$$

where η_m is a component known to the retailers but unknown to us. I assume that η_m has a discrete support $\{-\rho, \rho\}$; this term captures unobservable variation in permanent characteristics that make certain markets inherently profitable, and certain markets unprofitable. For this two-point support, I assume that with probability φ_1, $\eta_m = \rho$. Each retailer i's degree of uncertainty is captured by σ_i, which will ultimately generate differences in expectations among retailers that are informed, and retailers that are uninformed. Notice that if $\sigma_i = 0$, this model becomes a simple dynamic oligopoly game of entry/exit with unobserved heterogeneity.

4.3 Learning process

Through learning, retailers can resolve the uncertainty they face. In other words, learning should facilitate the process by which uninformed retailers become informed. With the framework above, I can build the process by which λ_{imt} evolves. Retailers have the prior $\lambda_0 \in (0,1)$, which is then updated in time t according to the following learning mechanism:

1. **Learning through entry:** Within a year of entering a market, a retailer becomes informed and resolves its uncertainty about the size of the market. Therefore, $\lambda_{imt} = 0$ if the retailer entered at time $t-1$. Furthermore, the retailer does not forget, so that $\lambda_{imt+s} = 0$ for all $s > 0$ if $\lambda_{imt} = 0$.

2. **Learning from others:** A potential entrant who has not previously entered (and left) the market already can learn from the observed past decisions of their informed rivals. The way in which the potential entrant updates the beliefs, λ_{imt}, is described in detail below.

To set up the process by which an uninformed potential entrant can learn from its peers, I first define the set of retailers that made informed decisions at time $t-1$:

$$J^*_{mt} = \{k : a_{kmt-2} = 1\}. \tag{5}$$

Note that each firm knows that every member in the set J^*_{mt} no longer faces uncertainty at period $t-1$. The vector of decisions among those that belong in the set of informed retailers J^*_{mt} at time $t-1$ is given by:

$$\boldsymbol{a}^*_{mt-1} = [a_{jmt-1} : j \in J^*_{mt}]. \tag{6}$$

With this notation in place and using Baye's rule, a potential entrant can then update its beliefs λ_{imt-1} using the following recursive equation:

$$\lambda_{imt} = \frac{\Pr(\boldsymbol{a}^*_{mt-1}|\boldsymbol{\omega}_m \neq \boldsymbol{0})\lambda_{imt-1}}{\Pr(\boldsymbol{a}^*_{mt-1}|\boldsymbol{\omega}_m \neq \boldsymbol{0})\lambda_{imt-1} + \Pr(\boldsymbol{a}^*_{mt-1}|\boldsymbol{\omega}_m = \boldsymbol{0})(1-\lambda_{imt-1})}. \tag{7}$$

Given the assumption of independent private information shocks, the conditional probability $\Pr(\boldsymbol{a}^*_{mt-1}|\cdot)$ is then defined as

$$\Pr(\boldsymbol{a}^*_{mt-1}|\cdot) = \prod\nolimits_{j \in J^*_{mt}} P_{jm}(\cdot)^{a_{jmt-1}} \cdot (1 - P_{jm}(\cdot))^{(1-a_{jmt-1})} \tag{8}$$

where $P_{jm}(\cdot) = \Pr(a_{jmt} = 1|\cdot)$. The probability $\Pr(\boldsymbol{a}^*_{mt-1}|\cdot)$ captures the information content associated with observed $\boldsymbol{a}^*_{mt-1}$, which is a vector of actions at period $t-1$ of these firms that belong to the set J^*_{mt}. With this learning process in place, it becomes clear what the components of the information set are:

$$\boldsymbol{\Omega}_{imt} = \{\boldsymbol{a}_{mt-2}, \boldsymbol{a}_{mt-1}, \boldsymbol{\lambda}_{mt-1}\}. \tag{9}$$

4.4 Markov Perfect Equilibrium (MPE)

The vector of payoff relevant state variables for firm is $(\boldsymbol{X}_{mt}, \varepsilon_{imt}, \eta)$. Here,

$$\boldsymbol{X}_{mt} = \{\boldsymbol{a}_{mt-2}, \boldsymbol{a}_{mt-1}, \boldsymbol{\lambda}_{mt-1}, \boldsymbol{Z}_{mt}\} \tag{10}$$

where $\boldsymbol{a}_{mt-2} = \{a_{imt-2}\}_i,\ \boldsymbol{a}_{mt-1} = \{a_{imt-1}\}_i,\ \boldsymbol{\lambda}_{mt-1} = \{\lambda_{imt}\}_i$ and $\boldsymbol{Z}_{mt}$ are exogenous market characteristics. An assumption I make regarding the equilibrium is that the strategy functions, $\{\varrho_i(\boldsymbol{X}_{mt}, \varepsilon_{imt}, \eta_m)\}_i$ depend on the state variables; hence, the equilibrium is Markov Perfect. Given this state, the equilibrium strategies can be written as

$$\varrho_i(\boldsymbol{X}_{mt}, \varepsilon_{imt}, \eta_m) = \arg \max_{a_{imt} \in \{0,1\}} E\left[\Pi^{\varrho}_{imt} + \beta\ V^{\varrho}_i(\boldsymbol{X}_{mt+1}, \varepsilon_{imt+1}, \eta_m) \mid \boldsymbol{X}_{mt},\ \varepsilon_{imt}, \eta_m\right] \tag{11}$$

where $V^{\varrho}_i(\boldsymbol{X}_{mt+1}, \varepsilon_{imt+1}, \eta_m)$ is the continuation value defined as

$$V^{\varrho}_i(\boldsymbol{X}_{mt}, \varepsilon_{imt}, \eta_m) = \max_{a_{imt} \in \{0,1\}} E\left[\Pi^{\varrho}_{imt} + \beta\ V^{\varrho}_i(\boldsymbol{X}_{mt+1}, \varepsilon_{imt+1}, \eta_m) \mid \boldsymbol{X}_{mt}, \varepsilon_{imt}, \eta_m\right]. \tag{12}$$

The one-shot payoffs Π^{ϱ}_{it} are evaluated at strategy ϱ. Integrating over the strategy function gives us

$$P_i(\boldsymbol{X}_{mt}, \eta_m) = \int_i \varrho(\boldsymbol{X}_{mt}, \varepsilon_{imt}, \eta_m) dG_i(\varepsilon_{imt}). \tag{13}$$

With this notation in place, the per-period expected profits are written as

$$E(\Pi^{\varrho}_{imt} \mid \boldsymbol{X}_{mt}, \varepsilon_{imt}, \eta_m) = \ a_{imt}[\Pi^{\mathbf{P}}_i(\boldsymbol{X}_{mt}, \eta_m) - \varepsilon_{imt}] \tag{14}$$

where $\Pi^{\mathbf{P}}_i(\boldsymbol{X}_{mt}, \eta_m)$ is defined in terms of expected market size and integrated strategies,

$$\Pi^{\mathbf{P}}_i(\boldsymbol{X}_{mt}, \eta_m) \equiv S_{mt}\theta_{1i} + \sum_{j \neq i} \theta_{2ij} P_j(\boldsymbol{X}_{mt}, \eta_m) - FC_i - (1 - a_{imt-1})EC_i + \eta_m + \lambda_{imt}\eta_m\sigma_i. \tag{15}$$

The expectation of the Bellman equation depends on the state vector, transition probability vector $F^{\boldsymbol{X},\boldsymbol{P}}_i(a_{imt}, \boldsymbol{X}_{mt}, \eta)$, and integrated value functions $\bar{\boldsymbol{V}}^{\mathbf{P}}_i$

$$E(V^{\varrho}_i(a_{imt}, \boldsymbol{X}_{mt+1}, \varepsilon_{imt+1}, \eta_m) | \boldsymbol{X}_{mt},\ \varepsilon_{imt}, \eta_m) \equiv F^{\boldsymbol{X},\boldsymbol{P}}_i(a_{imt}, \boldsymbol{X}_{mt}, \eta_m)' \, \bar{\boldsymbol{V}}^{\mathbf{P}}_i. \tag{16}$$

Here, each element of $\bar{\boldsymbol{V}}^{\mathbf{P}}_i$ is integrated over the future private information,

$$\bar{V}^{\mathbf{P}}_i(\boldsymbol{X}_{mt+1}, \eta_m) \equiv \int V^{\varrho}_i(\boldsymbol{X}_{mt+1}, \varepsilon_{imt+1}, \eta_m) dG(\varepsilon_{it+1}). \tag{17}$$

The best response function for firm i is now defined as

$$\varrho_i(\boldsymbol{X}_{mt}, \varepsilon_{imt}, \eta_m) = \begin{cases} 1 & \text{if } \Pi_i^{\mathbf{P}}(\boldsymbol{X}_{mt}, \eta_m) + \beta F_i^{\boldsymbol{X},\boldsymbol{P}}(1, \boldsymbol{X}_{mt}, \eta_m)' \bar{\boldsymbol{V}}_i^{\mathbf{P}} \\ & \geq \beta F_i^{\boldsymbol{X},\boldsymbol{P}}(0, \boldsymbol{X}_{mt}, \eta_m)' \bar{\boldsymbol{V}}_i^{\mathbf{P}} + \varepsilon_{imt} \\ \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

Consequently, the best response functions will satisfy

$$P_i(\boldsymbol{X}_{mt}, \eta_m) = G_i \left(\Pi_i^{\mathbf{P}}(\boldsymbol{X}_{mt}, \eta_m) + \beta [F_i^{\boldsymbol{X},\boldsymbol{P}}(1, \boldsymbol{X}_{mt}, \eta_m) - \ F_i^{\boldsymbol{X},\boldsymbol{P}}(0, \boldsymbol{X}_{mt}, \eta_m)]' \bar{\boldsymbol{V}}_i^{\mathbf{P}} \right). \tag{19}$$

Based on Aguirregabiria and Mira's (2007) representation lemma, the integrated values $\bar{\boldsymbol{V}}_i^{\mathbf{P}}$ can be expressed in terms of choice probabilities, which I describe in more detail in the Appendix.

5 Identification and estimation

5.1 Identification of the model

I allow for heterogeneity across retail chains. As I am able to observe each firm across many markets over the course of a many years, my data is rich enough to identify firm specific heterogeneity in entry and fixed costs, uncertainty, and strategic interactions. This heterogeneity is similar to that of Zhu and Singh (2009), as well as of Orhun (2012). The main difference is that their form of heterogeneity is defined by firm type, while mine is defined at the firm level. To some extent, one advantage of investigating the hamburger fast food industry is that the set of relevant players is quite small. If I allow a broader industry classification of quick service that also includes non-hamburger retailers, achieving such firm level heterogeneity comes at the cost of computation, as the state space of a dynamic game grows exponentially with the number of players.

There may in fact be multiple equilibria in the model. While multiple equilibria will affect computation of the model, it will not affect identification provided that we have appropriate exclusion restrictions (Tamer, 2003). In general, equilibrium uniqueness is not a requirement for identification of the model (Jovanovic, 1989). However, the main qualification for these results is that every observation in my sample pertains to the same equilibrium. Even if this qualification does not hold, the panel structure of my data allows for different equilibria across markets, as I observe the same players in all of the markets for 36 years.

When identifying the strategic interaction effects (θ_{2ij}), I encounter the same challenge as mentioned in previous studies of entry in marketing (Orhun, 2012; Vitorino, 2008; Zhu and Singh, 2011). The first source of bias is unobserved heterogeneity, which would result in the appearance of dampened competition or even complementarity between rival chains. In my model and estimation, I do allow for unobserved heterogeneity by introducing a market fixed effect, η_m. Most importantly, the introduction of dynamics aides in identification, as it provides an important exclusion restriction. For instance, a retail chain's incumbency status has a direct impact on its flow profits via the entry costs, but will only affect its rival through its best response probability $P_i(\boldsymbol{X}_{mt}, \eta_m)$.

However, the incumbency status only acts as an effective exclusion restriction if the chain is not active two periods earlier ($a_{imt-2} = 0$), or if the rival no longer faces any uncertainty about the market size ($a_{jmt-1} = 1$ or $\lambda_{jmt-1} = 0$). Otherwise, its decision to stay/exit will have a direct impact on the rival's payoff via the learning mechanism. Consequently, the parameters related to learning are confounded with the strategic interaction parameters. To separate out the parameters related to learning (λ_0, σ_i) from the strategic interaction effects, I need sufficient variation in a_{imt-2} and a_{imt-1}, given the functional form of the learning process as defined in my model.

Furthermore, (λ_0, σ_i) are also confounded with the market fixed effect η_m. In order to separately identify the parameters associated with learning from unobserved heterogeneity, I take advantage of one important source of variation generated by firm re-entry into markets. For example, consider a market in which a retail chain entered, left, and then re-entered. In my data, there are about 40 (out of 608) markets for which we see such behavior. The first time this chain entered, it most likely faced uncertainty. However, the second time it enters, the chain no longer faces uncertainty. In both cases, η_m is the same, but $(\lambda_{imt}, \sigma_i)$ enters through the payoff only in the first case. Furthermore, timing of its first entry helps identify the prior λ_0, as less weight is placed on the prior if the chain had more opportunities to learn from the past decisions of others.

Related to the issue of unobserved heterogeneity, there is likely an initial conditions problem in estimating this model, as some markets already have incumbents in the first year. In such cases, there could be a selection problem. To address this concern, I follow Arciacono and Miller's (2011) suggestion of using the first period observations to estimate the prior probability φ_1 of being in the a good market, where this prior probability is initialized using a flexible probit model.

In my model, the retail chains condition their strategies on the state $\boldsymbol{X}_{mt}$, which only contains information about the actions of competitors in the last two periods $(\boldsymbol{a}_{mt-2}, \boldsymbol{a}_{mt-1})$. It would appear as though the retailers were only learning based on these lagged decisions; therefore, my specification for their beliefs may not capture the full extent of their learned knowledge. However, the recursive structure of their learning process suggests otherwise. Note that their beliefs can be represented as a recursive relation $\lambda_{imt} = f(\lambda_{imt-1}, \boldsymbol{a}_{mt-2}, \boldsymbol{a}_{mt-1})$. If one solves this recursive relation, then their current period beliefs can actually be represented as $\lambda_{imt} = f(\lambda_0, \{\boldsymbol{a}_{mt-s}\}_{s>0})$.

Therefore, the inclusion of λ_{imt-1} as a state variable is a compact way of representing knowledge inferred from past decisions $\{\boldsymbol{a}_{mt-s}\}_{s>0}$. In other words, λ_{imt-1} is a sufficient statistic for $\{\boldsymbol{a}_{mt-s}\}_{s>0}$.

5.2 A simple DID specification test for learning

Using the framework set forth by my model, I can show the existence of a simple DID that can be adopted as an empirical test for the presence of learning using only the raw data patterns. This test will ultimately inform us as to the appropriateness of including a learning process in the dynamic entry/exit model. It however, will not tell us whether learning is actually causing the clustering behavior among fast food retailers. Therefore, this test only provides a first cut at detecting learning, and should primarily be used to motivate further analysis of the dynamic model I have presented via structural estimation and counterfactual analysis.

Consider the case in which there are only two retailers, i and j. Suppose that market m is one in which no retailer was ever active prior to $t-2$. Also, assume that this market is one in which rival j was active at time $t-2$. The focus here is to identify the learning effect for retailer i, as a consequence of j's past decision to stay as opposed to exit. Therefore, the observable states in $\boldsymbol{X}_{mt}$ collapse to (a_{imt-1}, a_{jmt-1}), since $a_{imt-2}=0$, $a_{jmt-2}=1$, and $\{a_{imt-s}=0\}_{s>2}$. Its one-shot payoff is $\Pi_i^{\mathbf{P}}(a_{imt-1}, a_{jmt-1}, \eta_m)$.

I deviate from the original model by making some additional assumptions. First, instead of a normal distribution for ε_{imt}, I assume that they follow a uniform distribution. Second, I set the discount rate for the retailers to be 0, so as to make them myopic. Therefore, one can then write a DID test in terms of the one-shot payoff as:

$$\begin{aligned}
\delta &= [\Pi_i^{\mathbf{P}}(0,1,\eta_m) - \Pi_i^{\mathbf{P}}(0,0,\eta_m)] - [\Pi_i^{\mathbf{P}}(1,1,\eta_m) - \Pi_i^{\mathbf{P}}(1,0,\eta_m)] \qquad (20)\\
&= \theta_{2ij}\{[P_j(0,1,\eta_m) - P_j(0,0,\eta_m)] - [P_j(1,1,\eta_m) - P_j(1,0,\eta_m)]\} + \sigma_i\eta_m[\lambda_{imt}(1) - \lambda_{imt}(0)]
\end{aligned}$$

The term $\lambda_{imt}(a)$ represents i's belief after observing $a_{jmt-1}=a$. With this expression, I now construct a null hypothesis for the scenario in which learning is not possible. As learning requires the presence of uncertainty, the DID test associated with the null hypothesis can be calculated by setting $\sigma_i = 0$. Consequently, the DID test of the expected one-shot payoffs is:

$$\begin{aligned}
\delta_0 &= \theta_{2ij}\{[P_j(0,1,\eta_m) - P_j(0,0,\eta_m)] - [P_j(1,1,\eta_m) - P_j(1,0,\eta_m)]\} \qquad (21)\\
&= \theta_{2ij}\theta_{2ji}\{[P_i(0,1,\eta_m) - P_i(0,0,\eta_m)] - [P_i(1,1,\eta_m) - P_i(1,0,\eta_m)]\}
\end{aligned}$$

Note that we can also write the best response probability as

$$
\begin{aligned}
P_i(a_{imt-1}, a_{jmt-1}, \eta_m) &= G_i(\Pi_i^{\mathbf{P}}(a_{imt-1}, a_{jmt-1}, \eta_m)) \qquad (22)\\
&= \Pi_i^{\mathbf{P}}(a_{imt-1}, a_{jmt-1}, \eta_m).
\end{aligned}
$$

Therefore, the DID test under the null hypothesis can also be represented as:

$$
\delta_0 = [P_i(0, 1, \eta_m) - P_i(0, 0, \eta_m)] - [P_i(1, 1, \eta_m) - P_i(1, 0, \eta_m)]. \qquad (23)
$$

The two expressions for the DID test under the null hypothesis equal each other *if and only if* $\delta_0 = 0$.

5.2.1 Is learning present in the fast food industry?

I illustrate this test by calculating the DID test for each chain-to-chain interaction based on a simple regression; to account for the market effects, I also condition on observed market characteristics and market fixed effects. Label $\Pi_i^{\mathbf{P}}(0, 1, \eta_m) = \Pi_i^{01}$, $\Pi_i^{\mathbf{P}}(0, 0, \eta_m) = \Pi_i^{00}$, $\Pi_i^{\mathbf{P}}(1, 1, \eta_m) = \Pi_i^{11}$, and $\Pi_i^{\mathbf{P}}(1, 0, \eta_m) = \Pi_i^{10}$. These objects can be estimated via the following regression:

$$
\begin{aligned}
E(a_{imt}|\boldsymbol{a}_{mt-1}, \boldsymbol{Z}_{mt}) &= (1 - a_{imt-1})(1 - a_{jmt-1})\Pi_i^{00} + (1 - a_{imt-1})a_{jmt-1}\Pi_i^{01} \qquad (24)\\
&\quad + a_{imt-1}(1 - a_{jmt-1})\Pi_i^{10} + a_{imt-1}a_{jmt-1}\Pi_i^{11} + \boldsymbol{Z}_{mt}\boldsymbol{\beta}_i + \eta_m.
\end{aligned}
$$

This linear regression is consistent with my assumption that the distribution of ε_{imt} is uniform. Here, η_m is unobserved market heterogeneity that I address using random effects. Note that I have to first rearrange the terms in order to obtain the double-difference δ. After expanding the terms and rearranging them, we get:

$$
E(a_{imt}|\boldsymbol{a}_{mt-1}, \boldsymbol{Z}_{mt}) = \Pi_i^{00} + (a_{imt-1} + a_{jmt-1})(\Pi_i^{10} - \Pi_i^{00}) - a_{imt-1}a_{jmt-1}\delta + \boldsymbol{Z}_{mt}\boldsymbol{\beta}_i + \eta_m \qquad (25)
$$

Therefore, δ can be estimated and its standard errors can be easily obtained as well. Table 9 presents the results from the DID regression, where column pertains to chain i, and row pertains to i's rival j. For these estimates, I cluster the standard errors at the market level. Note that for some chain-to-chain interactions, the DID is not equal to zero; in fact, some of these estimates are significant at a 10%-25% level. Note that as McDonald's exited only 2 markets in my entire sample,

Table 9: DID test for learning using fixed effects linear regression.

	A & W	Burger King	Harvey's	McDonald's	Wendy's
A & W		-0.001 (0.01)	0.04 (0.03)	-0.03 (0.02)	0.02 (0.01)
Burger King	-0.007 (0.02)		0.005 (0.02)	-0.02 (0.03)	-0.01 (0.02)
Harvey's	-0.03 (0.02)	0.02 (0.02)		-0.03 (0.02)	-0.02 (0.02)
McDonald's	N/A	N/A	N/A		N/A
Wendy's	-0.04 (0.03)	0.01 (0.04)	0.01 (0.03)	0.003 (0.008)	

I do not have enough variation to identify the DID effect McDonald's has on the other retailers. In general, there is also some heterogeneity in the DID across different retailers. This finding suggests that each retailer faces varying levels of *ex ante* uncertainty, as captured by my structural model. Ultimately, this model specification test provides reduced form evidence in favor of the presence of learning, and justifies the inclusion of uncertainty and learning in the structural model I estimate.

5.3 Estimation strategy

The parameters in my model are $\boldsymbol{\chi} = \{FC_i, EC_i, \theta_{1i}, \theta_{2ij}, \boldsymbol{\psi}, \sigma_i\}_{\forall i}$, λ_0, ρ, and φ_1. Therefore, conditional on $\boldsymbol{X}_{mt}$, and $\boldsymbol{\Upsilon} = \{\boldsymbol{\chi}, \lambda_0, \rho, \varphi_1\}$, the best response probability function $G_i(\cdot)$ is used to construct the pseudo-likelihood equation. To estimate the specification that incorporates a mixture distribution, I embed Arcidiacono and Miller's (2011) iterative Expectation-Maximization (EM) method with Aguirregabiria and Mira's (2007) Nested Pseudo Likelihood (NPL) procedure. A few additional steps are needed, as I outline in the Appendix. For notational simplicity, I use a subscript η to indicate the CCP associated with the unobserved state $\eta_m = \eta$. The criterion for optimization is:

$$Q(\boldsymbol{\Upsilon}, \boldsymbol{P}) = \sum_{i,m,t,\eta} \varphi_\eta LL[G_i(\boldsymbol{P}_{-i\eta}(\boldsymbol{X}_{mt}), \boldsymbol{X}_{mt}|\boldsymbol{\Upsilon})], \tag{26}$$

$$\begin{aligned} LL[G_i(\boldsymbol{P}_{-i\eta}(\boldsymbol{X}_{mt}), \boldsymbol{X}_{mt}|\boldsymbol{\Upsilon})] \quad = \quad & a_{imt} \log G_i(\boldsymbol{P}_{-i\eta}(\boldsymbol{X}_{mt}), \boldsymbol{X}_{mt}|\boldsymbol{\Upsilon}) \\ & +(1 - a_{imt}) \log[1 - G_i(\boldsymbol{P}_{-i\eta}(\boldsymbol{X}_{mt}), \boldsymbol{X}_{mt}|\boldsymbol{\Upsilon})]. \end{aligned} \tag{27}$$

This pseudo-likelihood is highly nonlinear in the prior probability λ_0, which make standard Newtonian optimization routines inefficient. Therefore, I consider an algorithm that essentially concentrates out $\boldsymbol{\chi}$, and then searches for λ_0 over a grid space. More details are provided in the Appendix. I base my method on the NPL as it does not require accurate non-parametric estimates for the initial CCPs $\boldsymbol{P}_0$ for consistency, while at the same time, being tractable. Moreover, the NPL estimates are more efficient than alternative two-step methods.[10]

Multiple equilibria would be a particular concern if I instead adopted a nested fixed point algorithm to estimate the game, as doing so would require explicitly solving the model for each maximum likelihood iteration. When using the NPL, the main concern are multiple NPL fixed points. One way to test whether the pseudo-likelihood yields multiple NPL fixed points is to initialize the NPL at randomly drawn first-stage CCPs. If the NPL fixed point and estimated parameters are the same for each initialization, then multiple NPL fixed points are unlikely to be an issue.

6 Main results

6.1 Summary of estimates

My structural estimates are summarized in Table 10. There is some heterogeneity in terms of each chain's cost structure. It is noteworthy is that McDonald's enjoys the highest brand value, as reflected in θ_{1MCD}. McDonald's high brand value in Canada should not be surprising, as it has always been the most recognized American brand in foreign countries. Also note that McDonald's has the highest entry costs. Part of its large entry costs could be a result of their extensive real estate research during *pro forma* analysis of prospective locations. Alternatively, their outlets may be the most expensive to build and/or make heavily advertised debuts. Fixed costs may be high for certain chains if they have a tendency to enter expensive markets.

Similar to Vitorino (2008), my estimates for strategic interaction suggest a potential for complementarity between certain chains (i.e., some $\theta_{2ij} > 0$), as the presence of a rival increases the expected payoff. In most cases though, the strategic interaction term has a negative sign, which suggests that most retailers treat one another as competitors. Among the retailers, it appears that McDonald's is the most sensitive about competition, while Burger King is the most sensitive to complementarity effects. This finding fits the anecdote that Burger King is often the one following McDonald's, and not the other way around. In fact, Toivanen and Waterson (2005) find that positive spillover effects in the U.K.'s fast food industry only benefit Burger King.

Most importantly, I find that chains face uncertainty, since for most of the retailers, the estimates for σ_i are non-trivial and statistically significant. As uncertainty is a prerequisite for learning, these results provide indirect evidence of learning. For most of the retailers, the degree of uncertainty

[10] Refer to Aguirregabiria and Mira (2010) for a comprehensive description of alternative methods.

Table 10: Structural estimation of dynamic entry/exit model.

	A & W	Burger King	Harvey's	McDonald's	Wendy's
Brand value (θ_{1ii})	0.08 (0.01)	0.01 (0.01)	0.01 (0.01)	0.03 (0.01)	0.02 (0.01)
vs A & W (θ_{2iAW})		0.05 (0.03)	0.1 (0.04)	-0.2 (0.03)	0.1 (0.03)
vs Burger King (θ_{2iBK})	-0.2 (0.04)		-0.04 (0.04)	-0.4 (0.03)	-0.05 (0.04)
vs Harvey's (θ_{2iHARV})	0.2 (0.03)	0.04 (0.03)		-0.2 (0.03)	0.09 (0.04)
vs McDonald's (θ_{2iMCD})	-0.3 (0.04)	-0.003 (0.04)	-0.09 (0.04)		-0.01 (0.03)
vs Wendy's (θ_{2iWEND})	0.05 (0.03)	0.03 (0.03)	0.07 (0.03)	0.07 (0.03)	
Fixed costs (FC_i)	-0.04 (0.04)	0.07 (0.04)	0.004 (0.05)	-0.4 (0.05)	0.07 (0.04)
Entry costs (EC_i)	0.1 (0.01)	0.03 (0.02)	0.08 (0.01)	0.02 (0.008)	-0.04 (0.01)
Degree of uncertainty (σ_i)	-0.2 (0.02)	0.03 (0.02)	-0.2 (0.02)	-0.3 (0.01)	-0.01 (0.02)
Prob. of uncertainty (λ_0)	0.20 (0.01)				
Good state parameter (ρ)	0.98 (0.14)				
Prob. of good state (φ_1)	0.45				

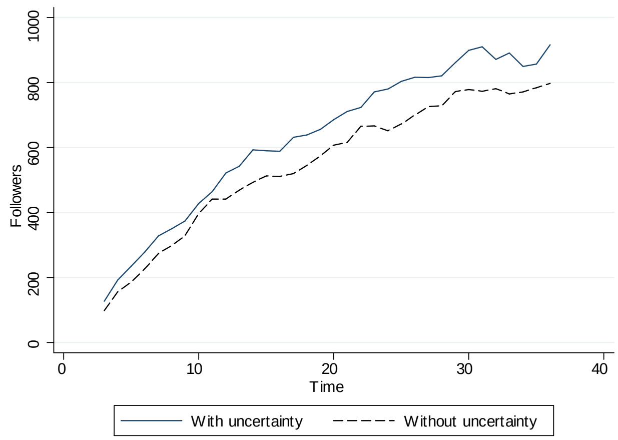
has a negative sign. What this means is that the retailers tend to bias downwards their beliefs about profitability in markets that are inherently good, and bias upwards their beliefs in those that are inherently bad. Also notice that the degree of uncertainty is different across the retailers; in particular, we see that σ_i is largest for A & W, Harvey's, and McDonald's, these retailers appear to be the most sensitive to uncertainty.

6.2 Can learning induce clustering?

The estimated structural model provides us an opportunity to look explicitly at the role of uncertainty in retail agglomeration. To investigate the impact of uncertainty on market outcomes, I compare the entry/exit decisions when uncertainty is present to when uncertainty is not present. One may interpret a counterfactual reduction of uncertainty as the hypothetical event where the Canadian government releases to the public its (initially) confidential detailed data on restaurant sales (by category) from tax returns, or detailed information about market characteristics such as traffic lights. Such a policy is realistic, as many municipalities in Canada have adopted an open data initiative.

The objective of this analysis is to establish a link between uncertainty and retail clustering. Since uncertainty and learning are closely intertwined, such a link implies a connection between clustering and learning. Empirical analysis in the earlier sections has already shown us that the entry/exit patterns we see in the data are consistent with the story of learning; but such analysis does not actually show how uncertainty/learning will impact retail concentration, as uncertainty

Figure 4: The number of instances in which a retailer follows a rival incumbent into a market.



is a variable that cannot be directly measured with observed data. In general, clustering is likely generated by learning if we see that retailers have a greater tendency to enter markets that rival incumbents have done well in when uncertainty is present. By examining the impact of uncertainty on industry dynamics, I can determine whether such behavior consistent with herding can indeed be generated by learning.

The counterfactual scenario is implemented by setting $\sigma_i = 0$ for each retailer so as to calculate counterfactual probabilities by solving the model under this hypothetical regime. As there may be multiple equilibria, I select the one generated by solving the fixed point problem initialized with the equilibrium probabilities. These counterfactual probabilities are used to generate a new sequence of entry/exit decisions for each market under the scenario of no uncertainty. I then compare the actual configurations ($\sigma_i \neq 0$) with the counterfactual configurations ($\sigma_i = 0$).

The counterfactual analysis illustrates that clustering activity caused by uncertainty and learning may be related to a retailer's tendency to follow past rival entrants into the same markets. Figure 4 illustrates this phenomenon by analyzing the number of instances retailers enter markets in which incumbent rivals were successful in (i.e., markets for which incumbents survived), under the scenarios with and without uncertainty. We see that when uncertainty is present, there are more instances in which such herding behavior occurs; and it is this difference that gives us a clue about the impact that learning has on clustering. Initially, the difference between the curves is quite small, but becomes larger over time. This pattern is consistent with learning, as informational spillovers become less noisy and more reliable as potential entrants take in more observational data via their rivals' past actions. Furthermore, retailers may have an incentive to delay their entry

Table 11: Average number of years before first entering a market.

	With uncertainty	Without uncertainty
A & W	5.0	4.0
Burger King	3.3	4.5
Harvey's	3.3	8.2
McDonald's	7.7	5.8
Wendy's	11.7	11.9

into markets so as to avoid being the first entrants into a market, whereby being first yields no informational spillover that they can get a free-ride off of (Chamley, 2004). Strategic delay would ultimately generate the pattern in Figure 4 where the herding behavior is more pronounced in the latter years, as the option value of delay falls.

Table 11 confirms that some chains may indeed be delaying their entry strategically in light of informational spillovers. When the average number of years it takes each retailer to enter a market is calculated for the industries with and without uncertainty, we see that A & W, and McDonald's take a disproportionately longer time to enter the market when uncertainty is present. In particular, A & W on average enters a market 1 year later when uncertainty is present, while McDonald's enters a market nearly 2 years later when uncertainty is present.

Not only does this counterfactual experiment demonstrate increased herding into good markets, it shows that retailers have a greater tendency to avoid markets that others have failed in when uncertainty is present as Figure 5 shows. When uncertainty is present, potential entrants react to observed exit by staying away from such markets. Ultimately, learning induces retailers to cluster via these two forces, the first drawing retailers into good markets, and the second, repelling retailers away from bad markets.

I now illustrate a trade-off that incumbents have to make in light of information externalities. The traditional thought in industrial organization stipulates a benefit of entering a market early due to entry deterrence. However, if one introduces learning from others to the discussion, then the benefit from entry deterrence is less clear. While entry deterrence prevents future competitors from entering, a retailer that enters early on takes on risk that may have been resolved via rational herding. Figure 6 illustrates such patterns, where it graphs the number of cases in which an incumbent that deters entry exits the market for the two industry scenarios. There are more of these cases when retailers face informational spillovers, which suggests some cost associated with entry deterrence. Ultimately, such costs may make herding more attractive than entering early.

Figure 5: The number of instances in which a retailer avoids a market others failed in.

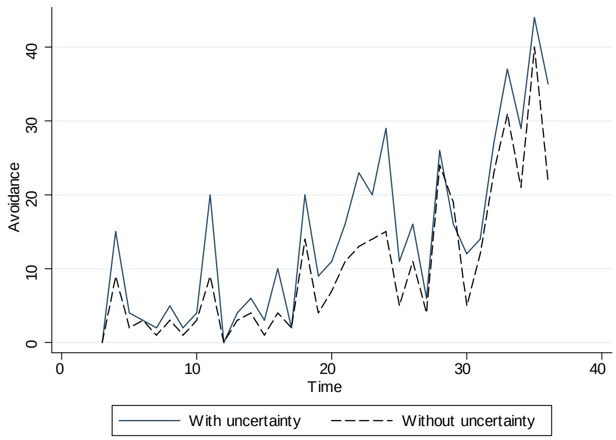
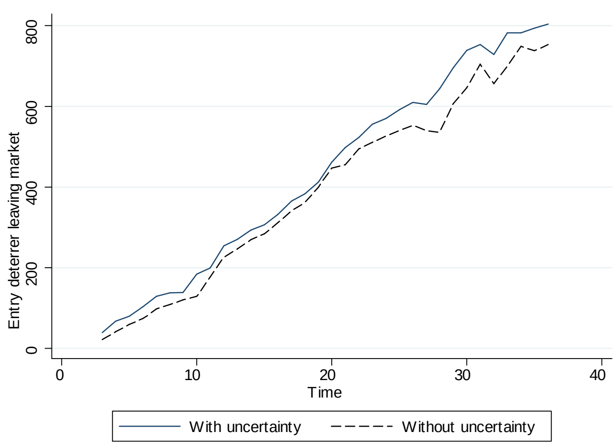


Figure 6: Number of instances in which entry deterrer exits a market.



7 Concluding remarks

The primary objective of my paper has been to understand an overlooked force behind clustering of retail chains, a challenging problem that current research in industrial organization, marketing, and urban economics has not yet explored. Using unique data with rich time and geographic variation from Canada's fast food industry, I develop a new oligopoly model of entry and exit that accounts for learning and unobserved heterogeneity. Using the model, I derive a simple DID test for learning, that when applied to my data, shows that learning is present in the fast food industry. Through counterfactual analysis of an estimated model, I show that an industry facing uncertainty and learning is more clustered than an industry facing no uncertainty and learning, thereby showing a connection between learning and agglomeration.

In future work, researchers may wish to consider that firms can potentially learn about profitability through their own experience in similar or neighboring markets. For example, a retail chain may learn through its past experience that low income markets are better than high income markets for generating demand if low income households have a greater propensity to consume unhealthy and salty food. Such experiences should then induce the chain to focus primarily on these markets in the future. My analysis has abstracted away from such learning behavior. However, it may be worthwhile considering this extension for future work as doing so can introduce rich heterogeneity in the *ex ante* beliefs that can ultimately be identified by data, when information about realized revenue is not available. With such a model, one can determine which *types* of markets are riskier than others. Such insight would especially be useful if retail managers have limited resources for conducting real estate research across markets, and wish to allocate their local headquarters optimally.

Finally, the DID (regression) test for learning I present need not be restricted to the fast food industry. It could in principle be applied to a more general class of social interaction models. For example, this test could be used to determine whether learning from peers is present in the adoption of new technologies, or in the consumption of new experience goods (i.e., word-of-mouth). The general strategy for this strand of empirical research is to first identify credibly a peer effect, and then run a series of *ad hoc* falsification tests that suggest that these peer effects are most likely driven by learning.

References

[1] Aguirregabiria, V. and Mira, P. (2007). Sequential Estimation of Dynamic Discrete Games. *Econometrica* 75.

[2] Aguirregabiria, V. and Mira, P. (2010). Dynamic Discrete Choice Models: A Survey. *Journal of Econometrics* 156, 38-67.

[3] Arcidiacono, P. and Miller, R. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica* 79. 1823-1867.

[4] Beamish, P., Jung, J. and Kim, H-H. (2003). American Fast Food in Korea. *Richard Ivey School of Business Foundation*, March 6, 2003.

[5] Bell, D. and Shelman, M. (2011). KFC's Radical Approach to China. *Harvard Business Review*, November 2011.

[6] Bennet, D., Chiang, C-F, and Malani, A. (2011). Learning During a Crisis: the SARS Epidemic in Taiwan. NBER Working paper.

[7] Blair, R-D. and Lafontaine, F. (2004). *The Economics of Franchising*. Cambridge University Press: New York.

[8] Bresnahan, T. and Reiss, P. (1991). Entry and competition in concentrated markets. *Journal of Political Economy* 99.

[9] Buera, F., Monge-Naranjo, A., and Primiceri, G. (2010). Learning the Wealth of Nations. *Econometrica* forthcoming.

[10] Caplin, A. and Leahy, J. (1998). Miracle on Sixth Avenue: Information Externalities and Search. *The Economic Journal* 108.

[11] Chamley, C. (2004). *Rational Heards: Economic models of Social Learning*. Cambridge University Press: Cambridge.

[12] Chamley, C. and Gale, D. (1994). Information Revelation and Strategic Delay in a Model of Investment. *Econometrica* 62.

[13] Chevalier, J. and Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research* 63, 345-354.

[14] Chi, F. and Yang, N. (2010). Twitter Adoption in Congress. *Review of Network Economics* 10.

[15] Conley, T. and Udry, C. (2010). Learning about a New Technology: Pineapple in Ghana. *American Economic Review* 100, 35-69.

[16] Damar, H-E. (2009). Why Do Payday Lenders Enter Local Markets? Evidence from Oregon. *Review of Industrial Organization* 34, 173-191.

[17] Datta, S. and Sudhir, K. (2011). The Agglomeration-Differentiation Tradeoff in Spatial Location Choice. Working paper.

[18] Deutsch, C. (1993). Commercial Property: Ladies' Mile; From the Carriage Trade to the Category Killers. *New York Times*, June 23, 1993.

[19] Dubé, J-P., Hitsch, G., and Rossi, P. (2010). State Dependence and Alternative Explanations for Consumer Inertia. *RAND Journal of Economics* forthcoming.

[20] Dunne, T., Klimek, S., Roberts, M. and Xu, Y. (2009). Entry, Exit, and the Determinants of Market Structure. NBER Working paper.

[21] Eckert, A. and West, D. (2008). Firm Survival and Chain Growth in a Privatized Retail Liquor Store Industry. *Review of Industrial Organization* 32.

[22] Ellickson, P. Houghton, S. and Timmins, C. (2010). Estimating Network Economies in Retail Chains: A Revealed Preference Approach. NBER Working paper.

[23] Ellison, G., Glaeser, E. and Kerr, W. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review* 100, 1195-1213.

[24] Eppli, M. and Benjamin, J. (1994). The Evolution of Shopping Center Research: A Review and Analysis. *Journal of Real Estate Research* 9.

[25] Feinberg, R. (2008). Explaining the Credit Union Entry Decision, and Implications for Performance. *Review of Industrial Organization* 33, 81-91.

[26] Goolsbee, A. and Klenow, P. (2002). Evidence on Learning and Network Externalities in the Diffusion of Home Computers. *Journal of Law and Economics* 45, 317-43.

[27] Holmes, E. (2010). U.S. Apparel Retailers Map an Expansion to the North. *Wall Street Journal*, March 29, 2010.

[28] Holmes, T. (2011). The Diffusion of Wal-Mart and the Economies of Density. *Econometrica* 79, 253-302.

[29] Jia, P. (2008). What happens when Wal-Mart comes to town: an empirical analysis of the discount retailing industry. *Econometrica* 76.

[30] Jovanovic, B. (1982). Selection and the Evolution of Industry. *Econometrica* 50.

[31] Jovanovic, B. (1989). Observable Implications of Models with Multiple Equilibria. *Econometrica* 57, 1431-1437.

[32] Kim, S-H. (2009). Predatory reputation in US airline markets. *International Journal of Industrial Organization* 27.

[33] Knight, B. and Schiff, N. (2007). Momentum and Social Learning in Presidential Primaries. *Journal of Political Economy* forthcoming.

[34] Kotler, P. and Keller, K-L. (2004). *Marketing Management*. Pearson Canada: Toronto.

[35] Konishi, H. Concentration of competing retail stores. *Journal of Urban Economics* 58, 488-512.

[36] Kosová, R. and Lafontaine, F. (2010). Survival and Growth in Retail and Service Industries: Evidence from Franchised Chains. *The Journal of Industrial Economics* 58.

[37] Lawrence, Requejo, and Graham. (2011). Health consciousness of consumers. *Harvard Business Review Blog*, May 19, 2011

[38] Love, J. (1995). *McDonald's: Behind the Arches*. Bantam Books: New York.

[39] Moon, Y. (2003). In-N-Out Burger. *Harvard Business School* case study.

[40] Moretti, E. (2010). Social Learning and Peer Effects in Consumption: Evidence from Movie Sales. *The Review of Economic Studies* forthcoming.

[41] Nishida, M. (2008). Estimating a model of strategic store-network choice. NET Institute working paper.

[42] Orhun, A-Y. (2012). Spatial Differentiation in the Supermarket Industry: The Role of Common Information. Working paper.

[43] Rob, R. (1991). Learning and Capacity Expansion under Demand Uncertainty. *The Review of Economic Studies* 58, 655-75.

[44] Schlosser, E. (2001). *Fast Food Nation*. HarperCollins: New York.

[45] Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics* 37.

[46] Shen, Q. (2010). A Dynamic Model of Entry and Exit in a Growing Industry. Working paper.

[47] Shen, Q. and Liu, H. (2012). Dynamic Learning and Exit in Competitive Markets. Working paper.

[48] Shen, Q. and Xiao, P. (2011). McDonald's and KFC in China: Competitors or Companions. Working paper.

[49] Smith, A. (2006). *Encyclopedia of Junk Food and Fast Food*. Greenwood Press: Westport.

[50] Strauss, M. (2011). U.S. retail invasion gathers pace. *The Globe and Mail*, January 21, 2011.

[51] Suzuki, J. (2010). Land Use Regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry. Working paper.

[52] Tamer, E. (2003). Incomplete Simultaneous Discrete Response Model with Multiple Equilibria. *Review of Economic Studies* 70, 147-167.

[53] Thomadsen, R. (2007). Product Positioning and Competition: The Role of Location in the Fast Food Industry. *Marketing Science* 26, 792-804.

[54] Thomadsen, R. (2010). Seeking an Aggressive Competitor: How Product Line Expansion Can Increase All Firms' Profits. Working paper.

[55] Toivanen, O. and Waterson, M. (2005). Market structure and entry: where's the beef? *RAND Journal of Economics* 36.

[56] Toivanen, O. and Waterson, M. (2011). Retail Chain Expansion: The Early Years of McDonald's in Great Britain. Working paper.

[57] Tucker, C. and Zhang, J. (2010). Growing Two-Sided Networks by Advertising the User Base: A Field Experiment. *Marketing Science* 29, 805-814.

[58] Varela, M. (2010). The Costs of Growth: Estimating Entry Costs with Endogenous Growth Rates. Working paper.

[59] Vitorino, M-A. (2008). Empirical Entry Games with Complementarities: An Application to the Shopping Center Industry. *Journal of Marketing Research* forthcoming.

[60] Yang, N. (2010). Burger King & McDonald's: Where's the Spillover? *International Journal of the Economics of Business* forthcoming.

[61] Zhang, J. (2010). The Sound of Silence: Observational Learning in the U.S. Kidney Market. *Marketing Science* 29, 315-355.

[62] Zhu, T. and Singh, V. (2009). Market Structure and Competition in the Retail Discount Industry. *Journal of Marketing Research* 46,453-466.

[63] Zhu, T. and Singh, V. (2009). Spatial competition with endogenous location choices: An application to discount retailing. *Quantitative Marketing and Economics* 7, 660-667.

[64] Zhu, T., Singh, V. and Dukes, A. (2011). Local competition, entry, and agglomeration. *Quantitative Marketing and Economics* forthcoming.

8 Appendix

8.1 Details about how variables are imputed

I impute the population in 1999 using the inferred exponential population growth rate between 1996 and 2001, and the population in 1990 using the exponential growth rate between 1991 and 1996. Observations before 1986 are imputed using a convex combination of the national growth rate and the growth rate pertaining to 1986 to 1991. I place a greatest weight on the annual national growth rate for years closest to 1970, and greatest weight on the 1986-1991 growth rate for years approaching 1986. I am also able to obtain the geographic area (in sq km) for each FSA from the Census of Canada. These values are later used to calculate the population density for each FSA market.

I impute income and property value in a similar manner as population. The difference is that for the years before 1986, I use a convex combination of the national inflation rate and the rate of return pertaining to 1986 to 1991. Because the proportion of residents who work in/out of an FSA market was not available for each Census, I use the information available for 2006.

8.2 Applying Aguirregabiria and Mira's (2007) representation lemma

I will now demonstrate how the MPE can be expressed using only the conditional choice probabilities, states, and model primitives. As before, $\boldsymbol{X}_{mt}$ denotes the state. We can express the specific values associated with being active and not as the following:

$$v_i(1, \boldsymbol{X}_{mt}, \eta) \equiv \Pi_i^{\mathbf{P}}(\boldsymbol{X}_{mt}, \eta) + \beta F_i^{\boldsymbol{X},\boldsymbol{P}}(1, \boldsymbol{X}_{mt}, \eta)' \bar{\boldsymbol{V}}_i^{\mathbf{P}}$$

$$v_i(0, \boldsymbol{X}_{mt}, \eta) \equiv \beta F_i^{\boldsymbol{X},\boldsymbol{P}}(0, \boldsymbol{X}_{mt}, \eta)' \bar{\boldsymbol{V}}_i^{\mathbf{P}}$$

where $F_i^{\boldsymbol{X},\boldsymbol{P}}(1, \boldsymbol{X}_{mt}, \eta)$ and $F_i^{\boldsymbol{X},\boldsymbol{P}}(0, \boldsymbol{X}_{mt}, \eta)$ are transition probability vectors, and $\bar{\boldsymbol{V}}_i^{\mathbf{P}}$ is a vector of integrated values across all possible states. Because the decision variable is discrete, we can write the integrated value as

$$\begin{aligned}
\bar{V}_i^{\mathbf{P}}(\boldsymbol{X}_{mt}, \eta) &\equiv P_i(\boldsymbol{X}_{mt}, \eta) v_i(1, \boldsymbol{X}_{mt}, \eta) + (1 - P_i(\boldsymbol{X}_{mt}, \eta)) v_i(0, \boldsymbol{X}_{mt}, \eta) + e_{imt}^{\boldsymbol{P}} \\
&= P_i(\boldsymbol{X}_{mt}, \eta)[\Pi_i^{\mathbf{P}}(\boldsymbol{X}_{mt}, \eta) + \beta F_i^{\boldsymbol{X},\boldsymbol{P}}(1, \boldsymbol{X}_{mt}, \eta)' \bar{\boldsymbol{V}}_i^{\mathbf{P}}] \\
&\quad + (1 - P_i(\boldsymbol{X}_{mt}, \eta))[\beta F_i^{\boldsymbol{X},\boldsymbol{P}}(0, \boldsymbol{X}_{mt}, \eta)' \bar{\boldsymbol{V}}_i^{\mathbf{P}}]
\end{aligned}$$

where $e_{imt}^{\boldsymbol{P}} = \phi(\Phi^{-1}(P_i(\boldsymbol{X}_t, \eta, \kappa))$, and $e_{imt}^{\boldsymbol{P}}$ is derived using the assumption that ε_{imt} has an iid normal distribution. The integrated values can be stacked into a vector across the states.

$$\bar{\boldsymbol{V}}_i^{\mathbf{P}} = \boldsymbol{P}_i[\boldsymbol{\Pi}_i^{\mathbf{P}} + \beta \boldsymbol{F}_i^{\boldsymbol{X},\boldsymbol{P}}(1)' \bar{\boldsymbol{V}}_i^{\mathbf{P}}] + (\boldsymbol{I} - \boldsymbol{P}_i)[\beta \boldsymbol{F}_i^{\boldsymbol{X},\boldsymbol{P}}(0)' \bar{\boldsymbol{V}}_i^{\mathbf{P}}] + \boldsymbol{e}_i^{\boldsymbol{P}}.$$

The term $\boldsymbol{P}_i$ is a stacked vector of conditional choice probabilities for i across all states. We can then obtain the integrated values with the following expression

$$\bar{\boldsymbol{V}}_i^{\mathbf{P}} = [\boldsymbol{I} - \beta \boldsymbol{F}_i^{\boldsymbol{X},\boldsymbol{P}}]^{-1} \{\boldsymbol{P}_i \boldsymbol{\Pi}_i^{\mathbf{P}} + \boldsymbol{e}_i^{\boldsymbol{P}}\}$$

$$\boldsymbol{F}_i^{\boldsymbol{X},\boldsymbol{P}} = \boldsymbol{P}_i \boldsymbol{F}_i^{\boldsymbol{X},\boldsymbol{P}}(1) + (\boldsymbol{I} - \boldsymbol{P}_i) \boldsymbol{F}_i^{\boldsymbol{X},\boldsymbol{P}}(0).$$

Using $\bar{\boldsymbol{V}}_i^{\mathbf{P}}$ defined above, we can therefore express the MPE above using only choice probabilities.

8.3 Details about the estimation procedure

The estimation algorithm can be described as follows:

1. Generate a grid of possible values for $\lambda_0^{(g)} \in [0, 1]$.

2. Estimate non-parametrically the initial CCP vector $\hat{\boldsymbol{P}}_\eta^{0,(g)}$. Alternatively, draw them randomly from a uniform distribution.

3. As in Arcidiacono and Miller (2011), initialize φ_η^0 at the predicted probability from a fitted probit model of entry using the first year's worth of data.

4. Given $\boldsymbol{X}_{mt}$, $\hat{\boldsymbol{P}}_\eta^{0,(g)}$, and $\lambda_0^{(g)}$, generate a sequence of posterior beliefs for each firm and market $\{\hat{\lambda}_{\eta mt}^{0,(g)}\}_{\forall \eta, m, t}$.

5. Given $\boldsymbol{X}_{mt}$, $\hat{\boldsymbol{P}}_\eta^{0,(g)}$, $\lambda_0^{(g)}$, and $\{\hat{\lambda}_{\eta mt}^{0,(g)}\}_{\forall \eta, m, t}$, compute:

$$q_m^{0,(g)} = \frac{\varphi_{\eta\kappa}^0 \prod_t LL[G_i(\boldsymbol{P}_{-i}(\boldsymbol{X}_{mt}, \eta_G), \boldsymbol{X}_{mt} | \boldsymbol{\chi})]}{\sum_{\eta'} \varphi_{\eta'}^0 \prod_t LL[G_i(\boldsymbol{P}_{-i}(\boldsymbol{X}_{mt}, \eta'), \boldsymbol{X}_{mt} | \boldsymbol{\chi})]}. \tag{28}$$

6. Use $q_m^{0,(g)}$ to calculate φ_η^1 according to:

$$\varphi_\eta^{1,(g)} = \frac{\sum_m q_m^{0,(g)}}{M}. \tag{29}$$

7. Given $\boldsymbol{X}_{mt}$, $\hat{\boldsymbol{P}}_\eta^{0,(g)}$, $\lambda_0^{(g)}$, and $\{\hat{\lambda}_{\eta mt}^{0,(g)}\}_{\forall \eta, m, t}$, find

$$\hat{\boldsymbol{\chi}}^{0,(g)} = \arg\max_{\boldsymbol{\chi}} Q(\{\boldsymbol{\chi}, \lambda_0^{(g)}\}, \hat{\boldsymbol{P}}_\eta^{0,(g)} | \boldsymbol{X}_{mt}, \{\hat{\lambda}_{\eta imt}^{0,(g)}\}_{\forall \eta, i, m, t}). \tag{30}$$

8. Update $\hat{\boldsymbol{P}}_\eta^{0,(g)}$ using $\hat{\boldsymbol{P}}_\eta^{1,(g)} = \{G_i(\hat{\boldsymbol{P}}_{-i}^{0,(g)}(\boldsymbol{X}_{mt}, \eta), \boldsymbol{X}_{mt} | \boldsymbol{X}_{mt}, \hat{\boldsymbol{\chi}}^{0,(g)}, \lambda_0^{(g)}, \{\hat{\lambda}_{\eta mt}^{0,(g)}\}_{\forall \eta, m, t})\}_{\forall m, t}$.

9. Repeat steps 2 to 5 until $\left\| \hat{\boldsymbol{P}}_\eta^{k+1,(g)} - \hat{\boldsymbol{P}}_\eta^{k,(g)} \right\|$ and $\left\| \varphi_\eta^{k+1,(g)} - \varphi_\eta^{k,(g)} \right\|$ are close to zero,[11] where k is equal to the number of iterations. Once convergence is reach, we have $\hat{\boldsymbol{P}}_\eta^{NPL,(g)}$, $\hat{\boldsymbol{\chi}}^{NPL,(g)}$, and $\hat{\varphi}_\eta^{(g)}$.

10. Do steps 2 to 10 for each possible value for $\lambda_0^{(g)}$, and then choose $\lambda_0^{(g)}$, $\hat{\varphi}_\eta^{(g)}$ and $\hat{\boldsymbol{\chi}}^{NPL,(g)}$ that is associated with the highest pseudo-likelihood function.

[11] I define the tolerance level to be 10^{-8} for convergence in both the NPL and likelihood maximization procedures.