Working Paper #11-16

October 2011

An Empirical Model of Industry Dynamics with Common Uncertainty and Learning from the Actions of Competitors

Nathan Yang
University of Toronto

An Empirical Model of Industry Dynamics with Common Uncertainty and Learning from the Actions of Competitors*

Nathan Yang†
University of Toronto

October 24, 2011

Abstract

This paper advances our collective knowledge about the role of learning in retail agglomeration. Uncertainty about new markets provides an opportunity for sequential learning, where one firm's past entry decisions signal to others the potential profitability of risky markets. The setting is Canada's hamburger fast food industry from its early days in 1970 to 2005, for which simple analysis of my unique data reveals empirical patterns pointing towards retail agglomeration. The notion that uninformed potential entrants have an incentive to learn, but not informed incumbents, motivates an intuitive double-difference approach that separately identifies learning by exploiting differences in the way potential entrants and incumbents react to spillovers. This identification strategy confirms that information externalities are key drivers of agglomeration. Estimates from a dynamic oligopoly model of entry with information externalities provide further evidence of learning, as I show that common uncertainty matters. Counterfactual analysis reveals that an industry with uncertainty is initially less competitive than an industry with certainty, but catches up over time. Furthermore, there are many instances in which chains enter markets they would have avoided had they not faced uncertainty. Finally, consistent with the interpretation of uncertainty as an entry barrier, I find that chains place significant premiums on certainty at proportions beyond 2% of their total value from being monopolists.

Keywords: Agglomeration, commercial real estate investment, dynamic discrete choice game, entry and exit, investment delay, market structure, retail competition.

JEL: C73, D83, L13, L66, L81, R00.

*[**JOB MARKET PAPER**] I am thankful for supervision from Victor Aguirregabiria, and careful guidance from my thesis advisors, Avi Goldfarb, Lu Han, and Junichi Suzuki. Comments from Ron Borkovsky, Andrew Ching, Victor Couture, Rahul Deb, April Franco, Matthew Grennan, Masakazu Ishihara, Panle Jia, Mara Lederman, Joshua Lewis, Michael Luca, Robert McMillan, Pedro Mira, Jeffrey Prince, Nathan Schiff, William Strange, Otto Toivanen, Mauricio Varela, Maria Ana Vitorino, Kitty Wang, Mo Xiao, and Juanjuan Zhang are much appreciated. A sincere thank you to participants at the Toronto IO Lunch Seminar, Econometric Society North American Summer Meeting, and Yale School of Management CCI for helpful discussions. I gratefully acknowledge funding from Social Sciences and Humanities Research Council Canada and the NET Institute (www.netinst.org). Many thanks to Melissa Pannozzo (Harvey's Canada), Patricia Simiele (McDonald's Canada), and Mark White (Wendy's Canada) for expert insight. The data on Canada's fast food industry were collected with the aide of patient staff from the Toronto Reference Library.

†Email: `nathan.yang@utoronto.ca`.

1 Introduction

The notion of a symbiotic relationship between rival entities is hard to rationalize under traditional models of oligopoly competition with entry deterrence (Schmalensee, 1978). Yet clustering of (seemingly competitive) retail establishments is commonly observed, and has been well-documented by researchers: Burger King persistently opens new stores near existing McDonald's outlets (Toivanen and Waterson, 2005), and anchor stores tend to enter the same malls (Vitorino, 2008). Although such patterns provide convincing evidence of agglomeration, our knowledge about the causes underlying these patterns remains incomplete. Identifying the cause of agglomeration[1] is a challenging task due to the multitude of potential hypotheses. Unobserved heterogeneity and demand externalities are typical explanations for retail clustering. A nearby mall, local attraction, or highway exit can easily generate retail agglomeration among rivals (Thomadsen, 2007), as can restrictive retail zoning provisions (Datta and Sudhir, 2011) - both factors pointing to 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[2] concerns (Thomadsen, 2010; Zhu, Singh and Dukes, 2011).[3]

Despite the well-developed theoretical literature on social learning and learning-from-others,[4] empirical research on retail agglomeration has overlooked the idea that if managers face uncertainty about market profitability, then they may have an incentive to take advantage of any information that can possibly be revealed when an existing and informed chain decides to stay or exit a market. My objective is to establish the role of these information externalities in generating retail clustering, along with implications that uncertainty and such externalities have on an industry's evolution. From a manager's perspective, learning from the actions of others is potentially a profitable and cost-saving strategy, as retail chains spend considerable resources on real estate research. In particular, observing that a rival has failed in a particular market allows them to avoid sub-optimal choices and focus their attention on more promising investments.[5] Of interest to competition authorities,

[1] For a perspective from urban economics, refer to Rosenthal and Strange's (2004) survey and the recent study by Ellison, Glaeser, and Kerr (2010).

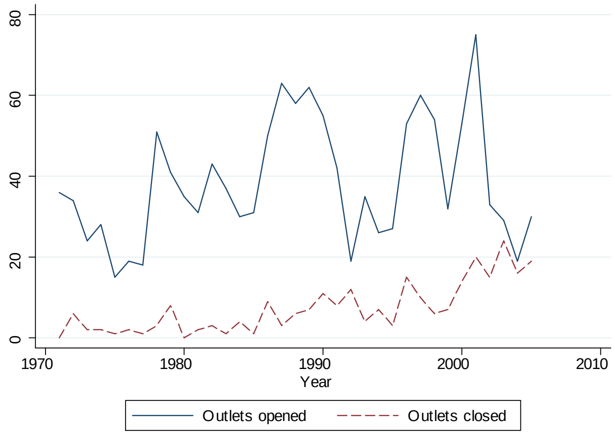
[2] For example, if a market has one McDonald's outlet, and one Burger King outlet, the entry of one additional McDonald's outlet can actually benefit both chains. The additional McDonald's outlet will induce McDonald's to price less competitively so as to avoid cannibalizing its original store's sales.

[3] On a related note, Sen, Shin and Sudhir (2011) find evidence of real monetary effects associated with demand externalities.

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

[5] The oft-cited newspaper article by Deutsch (1993) claims that the observable success of a Bed, Bath and Beyond outlet was a key factor that induced other retailers to set up shop in a New York neighborhood.

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



information externalities may even reduce the impact of *ex ante* uncertainty as an entry barrier.

This paper studies 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[6] around 1970 to 2005 (Section 2). As Canada is the first foreign market many of the American chains explored, it is likely that managers faced some uncertainty early on. Combined with (unusually) large time-series variation in both entry and exit (Figure 1),[7] these features make Canada's ever-evolving fast food industry a particularly attractive setting to study the relationship between learning and agglomeration.

Section 3 presents three empirical regularities. First, a chain's growth rate in a local market is positively associated with the past size of its rivals. Second, entry occurs earlier in local markets initially endowed with rival incumbents. And finally, 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 and the existence of spillovers.

Credible identification of information externalities is particularly challenging given perfectly valid alternative hypotheses involving demand externalities and unobserved heterogeneity. To guide

[6] Other studies that investigate empirical patterns in retail industry dynamics are Eckert and West (2008) and Kosová and Lafontaine (2010), both of which are motivated by the theoretical framework of Jovanovic (1982).

[7] Large relative to other retail industries, such as general merchandise and grocery. For instance, about 17% of all actions in my sample are decisions to exit a market. Compare this number with Beresteanu, Ellickson, and Misra's (2010) study about supermarkets that observes exit in about 4% of their observations.

my identification approach, I first present a dynamic oligopoly model of entry[8] that allows for common uncertainty and information externalities in Section 4.[9] The basic idea of the model is that *ex ante*, chains face uncertainty about market size and can at best obtain a forecast based on observable market characteristics. This uncertainty is resolved after entry, as incumbents will observe the true market size.[10] The decisions of incumbents will be made without uncertainty about market size, thereby giving rivals who have not yet entered an opportunity to learn from these observed stay/exit decisions.

Straightforward intuition consistent with the model suggests a semi-parametric double-difference decomposition that can be used to identify learning (Section 5).[11] The assumption that uncertainty is resolved upon entry will provide an incentive for uninformed potential entrants to learn, but not informed incumbents. Non-learning factors, such as unobserved heterogeneity or demand externalities, will affect both potential entrants and incumbents alike.[12] Therefore, a potential entrant's reaction to its rival's stay/exit decision will be driven by both learning and non-learning factors, while an incumbent's reaction to its rival's stay/exit decision will only be driven by non-learning factors. The difference in how a potential entrant and incumbent reacts to its rival's stay/exit decision will net out the non-learning effects, thereby isolating the contribution of learning. This approach demonstrates that information externalities are larger than the other effects; hence, learning is a key driver of retail agglomeration in Canada's hamburger fast food industry. Furthermore, these externalities affect each chain differently, which I explore in more detail with structural estimation.

I take advantage of structural estimation in Section 6 to show that the fast food chains do indeed face *ex ante* uncertainty, which is a necessary condition for learning.[13] Subsequent counterfactual analysis in Section 7 allows me to assess the implications of uncertainty for market structure and chain profitability. By simulating a counterfactual equilibrium for which the degree of *ex ante*

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

[9] For a summary of research about learning, refer to Ching, Erdem and Keane (2011).

[10] For example, a manager will find out through realized revenue reports the true profitability of markets.

[11] Other attempts at identifying social learning include: computer purchase decisions (Goolsbee and Klenow, 2002), educational product adoption (Forbes, 2009), employment adjustments (Guiso and Schivardi, 2007), entry into work among women (Fogli and Veldkamp, 2008), farming technology adoption (Conley and Udry, 2010), HIV/AIDS risk perception (Kohler, Behrman, and Watkins, 2007), home sales (Tucker, Zhang, and Zhu, 2009), internet adoption (Ward, 2010), kidney adoption (Zhang, 2010), macroeconomic policy choice (Buera, Monge-Naranjo, and Primiceri, 2010), momentum effects in sequential elections (Knight and Schiff, 2007), movie sales (Moretti, 2010), predatory behavior of incumbent airlines (Kim, 2009), ranking of college football teams (Stone and Zafar, 2011), SARS risks (Bennett, Chiang, and Malani, 2011), stock purchasing decisions (Ivkovic and Weisbenner, 2007), Twitter adoption among politicians (Chi and Yang, 2010), and word of mouth in online book sales (Chevalier and Mayzlin, 2006).

[12] For example, the additional traffic generated by the presence of a neighboring local attraction or rival benefits a chain, regardless of whether it just recently entered the market, or has been active for a number of years.

[13] Refer to the Appendix.

uncertainty is set to zero, I find that an industry is less competitive with uncertainty than without uncertainty during the first 20 years; however, we see a convergence towards the scenario with certainty over time. It has often been thought that first movers into markets are predisposed to creating entry barriers for future entrants (Bronnenberg, Dhar, and Dubé, 2009). But as I illustrate with this counterfactual simulation, when retail establishments face entry barriers related to uninsurable risk, the presence of incumbent rivals may reduce these barriers through the information externalities. Ultimately, the existence of these exploitable features naturally regulates markets and most likely contributes to the competitiveness of Canada's fast food industry (Figure 2). If firms are able to resolve their uncertainty through learning, then the presence of rival incumbents could actually be welfare-improving. Entering *ex post* unprofitable markets followed by exit is suboptimal, and my counterfactual analysis shows that uncertainty is a key driver for such market inefficiencies as an industry with uncertainty is associated with over 2% more entry/exit in comparison to an industry without uncertainty. However, these inefficiencies are less pronounced in the latter years as firms are less likely to enter markets they would have avoided under full information.

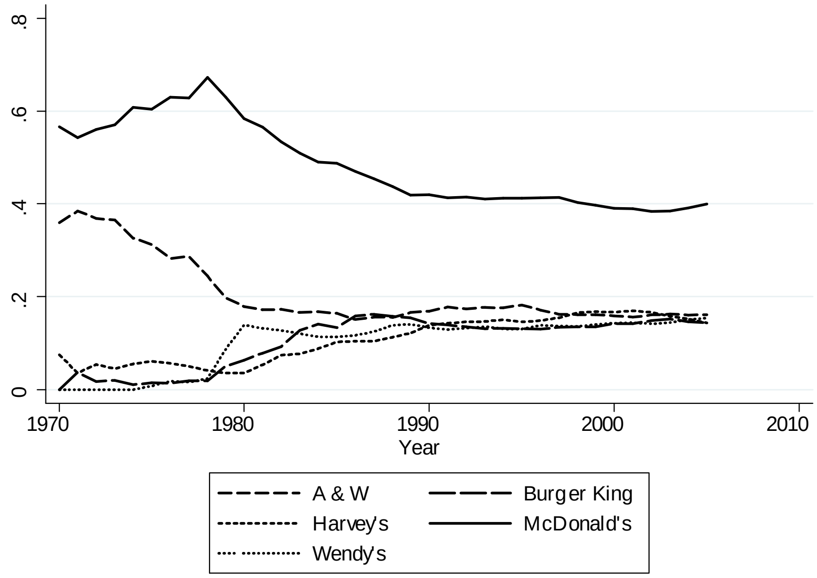
The counterfactual analysis also demonstrates that chains place a premium on certainty at levels exceeding their entry costs, providing further evidence that *ex ante* uncertainty is a significant entry barrier. In fact, these premiums constitute a non-negligible proportion of a monopolist's value in a local market of median size. Finally, variation in the way chains value a reduction in uncertainty could explain some of the variation in chains' reactions to information externalities as illustrated by the double-difference decomposition. In particular, the chain that reacts the least (most) to information externalities also places the least (most) value on uncertainty reduction.

2 Data

2.1 Canada's hamburger fast food industry

This study investigates local competition between 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, so, 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

Figure 2: Evolution of market shares in Canada's fast food industry.



another.[14]

Since 1970, Canada has become a very important foreign market for American retail chains. Canada provides American chains a real growth option,[15] 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 follow 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

[14] 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.

[15] 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.

most dominant forces in Canada's fast food industry today.

While franchisees have some freedom with respect to own-store pricing and promotions, the location decisions are made by the chain.[16] In particular, a typical fast food chain will have a real estate research team which is assigned to this task. During expansionary periods, this part of the company is quite large; so much so that each team member has a specific region that he overlooks.

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.[17] An FSA can be as large as a small city, or can be one of many subdivisions in a large metropolitan city.[18] The FSA regions I consider are those nested within Canada's Census Metropolitan Areas (CMAs). Loosely speaking, a CMA is a (major) city in Canada. My set of CMAs covers all of the provinces in Canada, although a large proportion of them are concentrated in the province of Ontario. Note that the FSAs are the largest subdivisions that I am able to find for my set of CMA's.[19] I find 608 FSA markets based on the cities used in my sample.

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[20] the missing years using the inferred population growth rates. Table 1 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.[21] Property value is used as a proxy for the cost of purchasing a location to house a fast food outlet.

[16] Franchisees can own multiple contracts with the fast food chains. The concentration of ownership has implications on the prices paid by consumers (Thomadsen, 2005).

[17] FSAs are on average 1.8 square miles in many Canadian cities, and thus, comparable to American Census Tracts. Note that these markets 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 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.

[18] Refer to the Appendix for a graphical example displaying the FSA regions for the city of Toronto.

[19] Population ranges from 44 to 89,696 people.

[20] 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.

[21] 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.

Past studies have shown that fast food chains prefer markets with poorer inhabitants.[22]

Table 1: 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

I supplement the Census data with the Small Area Retail Trade Estimators (SARTE). These 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. This counts all retail locations that belong to chains with at least 4 stores. SARTE is the most reliable dataset of retail sales at such a disaggregated level. However, its time series variation might not be reliable.[23] 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.[24] To rule out these markets, I exclude from my sample markets that have either zero retail sales/locations, population, or income.[25] After these inclusions, 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.

[22] Refer to Block, Scribner, and DeSalvo (2004) and Powell, Chaloupka, and Bao (2007).

[23] 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.

[24] This may especially be the case for small cities around the 1970s.

[25] 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.

2.3 Market structure data

I turned to archived phone books at the City of Toronto's Reference Library[26] 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.[27] 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.

Table 2 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 market-time observations have a chain operating more than 1 outlet. Table 3 shows that both entry and exit occur for most configurations. From this table, we see that exit is most likely to happen in markets with 4 or 5 chains active in the period before, while entry is most likely to happen with 0, 3 or 4 chains active in the period before. Finally, the chains in general differ in terms of their entry timing (Table 4). After tabulating the total number of markets each chain is a first entrant reveals that A & W and McDonald's typically enter first. Burger King, Harvey's, and Wendy's are more often than not followers into markets.

[26] This library has the most comprehensive collection of archived phone directories in Canada.

[27] In the event that Geocoder.ca was unable to find a postal code match corresponding to the desired address, I used either Google Maps, online store locators on the chains' websites, or inquired the chains directly by phone.

Table 2: Tabulation of market-time observations that contain 0, 1, ..., 9 outlets belonging to each of the chains. Note that these are not market structure configurations: each column should be read (and interpreted) independent of another column.

	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

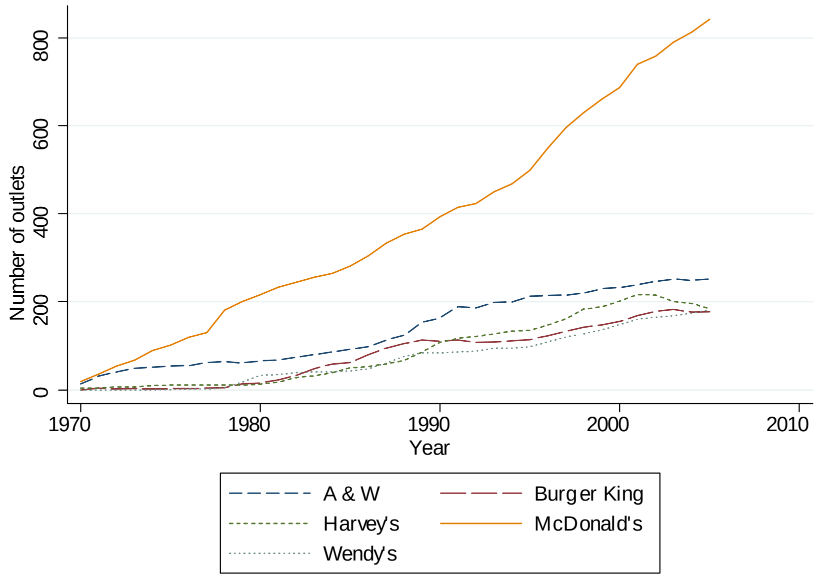
Table 3: Transition probabilities from X chains (row) to Y chains (column).

	0	1	2	3	4	5
0	93.32	6.46	0.21	0.01	0	0
1	1.06	93.87	4.69	0.33	0.04	0
2	0	2.48	90.93	6.30	0.26	0.04
3	0	0.07	3.13	90.54	6.05	0.21
4	0	0	0	5.23	90.97	3.80
5	0	0	0	0	6.08	93.92

Table 4: 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

Figure 3: Evolution of Canada's fast food industry.



3 Preliminary evidence on clustering and spillovers

3.1 Suggestive patterns[28]

We see from Figure 3 that the industry has experienced significant growth over the past few decades. This observation motivates me to look at what drives chain expansion. To measure the growth rate of each chain i in market m, I calculate market level growth as $Growth_{imt} = \log(Total_{imt+1} + 1) - \log(Total_{imt} + 1)$, where t is a time index. This definition of the growth rate allows for the fact that in many markets, $Total_{imt}$ or $Total_{imt+1}$ may be equal to zero[29]. In this specification, $Total_{jmt}$ is equal to the total number of outlets that j has in market m at time t, and $Total\_Other_{jmt}$ is the total number of outlets that j has in all other markets except m. Temporal and geographic variation of this measure allows me to look at the following panel data VAR model:

$$Growth_{imt} = \alpha_i + \boldsymbol{Z}_{mt}\boldsymbol{\beta}_i + \sum_{\forall j} \gamma_{1ij} \log(Total_{jmt} + 1) + \sum_{\forall j} \gamma_{2ij} \log(Total\_Other_{jmt} + 1) + \eta_m + \varepsilon_{imt}. \tag{1}$$

The coefficient γ_{1ii} captures the usual relationship between firm growth and own size within

[28] The Canadian fast food industry exhibits a lot of agglomeration. When I calculate the Ellison and Glaeser (1997) statistic using my data, the statistic, which lies between 0.01 to 0.08, is comparable to those pertaining to the automobile and computer industry, both of which are quintessential examples of localized industries.

[29] If for example, $Total_{imt} = Total_{imt+1} = 0$, then $Growth_{imt} = 0$.

market m, while γ_{1ij} captures the rarely analyzed relationship between firm growth and the size of others within market m. The vector $\boldsymbol{Z}_{mt}$ contains market specific information about the population, population density, income, and property value for a given city and year.

From Table 5, I would like to first point out that there is a negative relationship between firm growth and own size for each of the chains. This finding is consistent with Kosová and Lafontaine's (2010) analysis of franchised chain growth. Their explanation for this negative correlation is each chain's convergence to some equilibrium size. They attribute the boundedness of size to the fact that franchised chains often provide single products, which establishes a natural limit to "how much of a single product a firm produces."

Furthermore, there is a positive relationship between firm growth and the past size of others (in the same city): A & W's growth is positively associated with the size of Burger King and Wendy's; Burger King's growth is positively associated with Harvey's and Wendy's; McDonald's growth is positively associated with the size of A & W; and Wendy's growth is positively associated with the size of all of its rivals. Note that as the tests for first and second order serial correlation are unable to reject the null hypothesis of no serial correlation, the regression model is dynamically complete, and hence establishes *Granger causality.*

I now focus on patterns in entry timing. To investigate this relationship, I exploit the fact that my data contains a large amount of variation in terms of entry timing. The following regression is run using data on years of entry:

$$Year_{im} = \alpha + \gamma Initial_{im} + \zeta_i + \nu_m + \varepsilon_{im}. \tag{2}$$

where $i \in M_i$ where M_i is the set of markets that experienced entry by chain i. $Year_{im}$ defines the year in which entry occurred for the observation, and $Initial_{im}$ is the number of active rivals in 1970. I also include a chain fixed ζ_i and market fixed effect ν_m to address concerns of unobserved heterogeneity. Note that A & W entered 270 markets, Burger King entered 203 markets, Harvey's entered 232 markets, McDonald's entered 461 markets, and Wendy's entered 182 markets, which means that there is a total to 1348 instances of entry. The parameter γ captures the suggestive spillover effect on entry timing. With my data, the total number of rival chains in 1970 ranges from 0 to 3. The regressions suggest that having a large number of rival chains is correlated with early entry (Table 6).

Based on the observed patterns I have outlined in this section, one may conjecture the existence of externalities between chains. In the next section, I provide empirical evidence that indeed this may be the case.

Table 5: Relationship between market expansion growth and chain size over time with FSA level fixed effects. Estimation uses data at the chain-fsa-time level. Growth is calculated for each chain as $\log(Total_{imt+1}+1)-\log(Total_{imt}+1)$. I use the Arellano-Bond test for serial correlation in panel data for the 1st and 2nd order tests. This test statistic has an asymptotic Normal distribution. I include the test statistic along with its corresponding p-value in parentheses.

	(1) A & W	(2) Burger King	(3) Harvey's	(4) McDonald's	(5) Wendy's
log(A & W size same FSA)	-0.110***	0.00425	-0.00158	0.0145***	0.0125***
	(0.00330)	(0.00259)	(0.00301)	(0.00383)	(0.00243)
log(Burger King size same FSA)	0.0130**	-0.0968***	0.00224	-0.00213	0.00584
	(0.00416)	(0.00326)	(0.00378)	(0.00483)	(0.00305)
log(Harvey's size same FSA)	0.00317	0.00618*	-0.100***	-0.00151	0.0109***
	(0.00371)	(0.00291)	(0.00338)	(0.00431)	(0.00273)
log(McDonald's size same FSA)	0.00325	0.00290	0.00413	-0.108***	0.0118***
	(0.00277)	(0.00217)	(0.00252)	(0.00322)	(0.00203)
log(Wendy's size same FSA)	0.0251***	0.0104**	0.00434	0.00746	-0.0882***
	(0.00441)	(0.00346)	(0.00401)	(0.00512)	(0.00324)
log(A & W size other FSA)	-0.00265	-0.0391***	0.0149*	0.0120	-0.0113*
	(0.00752)	(0.00591)	(0.00685)	(0.00873)	(0.00553)
log(Burger King size other FSA)	0.00237	0.00440*	0.00273	-0.0101***	0.00178
	(0.00252)	(0.00198)	(0.00229)	(0.00292)	(0.00185)
log(Harvey's size other FSA)	0.00872**	0.0127***	0.0000182	0.0166***	0.000544
	(0.00324)	(0.00254)	(0.00295)	(0.00376)	(0.00238)
log(McDonald's size other FSA)	-0.0121*	0.0213***	-0.0163***	0.00580	0.0114**
	(0.00525)	(0.00412)	(0.00477)	(0.00609)	(0.00385)
log(Wendy's size other FSA)	0.00140	-0.00524**	0.00403	0.00157	-0.00199
	(0.00255)	(0.00200)	(0.00232)	(0.00296)	(0.00187)
log(Population)	-0.00120	-0.0297*	-0.0107	0.00295	-0.00106
	(0.0173)	(0.0136)	(0.0157)	(0.0201)	(0.0127)
log(Population density)	0.00347	0.0325*	0.0104	0.00264	0.00336
	(0.0185)	(0.0145)	(0.0168)	(0.0214)	(0.0136)
log(Income)	0.00417	-0.00395	-0.0171***	0.000898	-0.00110
	(0.00383)	(0.00300)	(0.00348)	(0.00444)	(0.00281)
log(Property value)	0.00106	0.000451	-0.00184	0.00334	-0.00108
	(0.00289)	(0.00227)	(0.00263)	(0.00335)	(0.00212)
Proportion work in same FSA	-0.0337	0.0290	0.0109	-0.0273	-0.0565
	(0.0834)	(0.0654)	(0.0758)	(0.0968)	(0.0612)
Constant	-0.00140	0.110	0.246***	-0.146	0.0370
	(0.0797)	(0.0625)	(0.0725)	(0.0925)	(0.0585)
Observations	20930	20930	20930	20930	20930
R^2	0.0540	0.0452	0.0457	0.0545	0.0376
Test for 1st-order autocorrelation	-0.62696 (0.5307)	-1.9385 (0.0526)	-0.78498 (0.4325)	-0.8418 (0.3999)	-1.3009 (0.1933)
Test for 2nd-order autocorrelation	0.70658 (0.4798)	0.91798 (0.3586)	0.33925 (0.7344)	0.9501 (0.3421)	-0.11496 (0.9085)

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

Table 6: Patterns in the timing of entry. The dependent variable is the year in which a chain first entered a market.

	(1) Year of entry	(2) Year of entry
Number of active rivals in 1970	-2.274***	-1.679***
	(0.161)	(0.377)
Burger King	1.974*	3.520**
	(0.766)	(1.145)
Harvey's	3.142***	2.654*
	(0.735)	(1.043)
McDonald's	-3.519***	-1.492
	(0.708)	(0.906)
Wendy's	3.101***	3.405**
	(0.821)	(1.157)
Number of active rivals in 1970 * Burger King		-1.056
		(0.557)
Number of active rivals in 1970 * Harvey's		0.383
		(0.537)
Number of active rivals in 1970 * McDonald's		-1.669***
		(0.501)
Number of active rivals in 1970 * Wendy's		-0.270
		(0.567)
Constant	1992.4***	1991.6***
	(0.599)	(0.724)
Observations	1348	1348
FSA fixed effects	No	Yes
R^2	0.1787	0.1917

Robust standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

3.2 Empirical evidence of spillover effects

The primary objective of this section is to establish the existence of spillover effects (or learning) using simple reduced form analysis. A main issue that we have to deal with in this section is controlling for unobserved heterogeneity of local markets so that we can be confident that our estimates of learning/spillover effects do not capture spuriously this market heterogeneity; this main finding is important as empirical studies that have produced positive correlations between entry decisions of seemingly rival entities have often blamed their "counter-intuitive" findings on unobserved heterogeneity.[30] I now argue that unobserved heterogeneity is not the sole explanation.

My goal is to verify a positive relationship between rivals' incumbency statuses, and one's own decision to enter/stay in a market. Establishing this relationship is akin to finding evidence of state dependence. The difference here is that not only your state, but also your rivals' states may matter. The model I wish to estimate for each fast food chain is thus

$$\Pr(a_{imt} = 1 | \boldsymbol{a}_{mt-1}, \boldsymbol{Z}_{mt}) = \Phi(\alpha_i + \boldsymbol{Z}_{mt}\boldsymbol{\beta}_i + \sum_{j \neq i} \gamma_{ij} a_{jmt-1} + \rho_i t + \eta_m + \varsigma_i t \cdot \eta_m) \tag{3}$$

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 so-called spillover effect is represented by γ_{ij} for all $i \neq j$; this is the effect I am interested in identifying. The time trend is captured by ρ. For the time being, I remain agnostic as to the interpretation of spillover so as to maintain generality. The main complication associated with estimating a model of this sort is the unobserved heterogeneity, captured by η_m, and its interaction with time.[31] Unobserved heterogeneity under the context of fast food competition may be interpreted as information that fast food developers have that is omitted in my data. Some examples would be information they obtain through proprietary market research companies or their own research teams. I estimate the market fixed effects directly using my 35 annual observations for each of the 608 FSA markets.

Table 7 provides the first set of evidence in favor of positive spillovers:[32] 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

[30] For example, Dunne, Limek, Roberts, and Xu (2009) employ this argument to explain why they find that the number of competitors appear to increase variable profits.

[31] I include this interaction as the unobserved heterogeneity may not be constant over 36 years.

[32] I get similar results using the Wooldridge (2005) extension of the random effects probit.

Table 7: Evidence of spillover effects in the chains' decision to be active in market. The estimates are obtained using the fixed effects probit estimator that includes time trends and interactions between time and FSA dummies.

	(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 spillover effects in the chains' decision to enter a market. The estimates are obtained using the fixed effects probit estimator that includes time trends and interactions between time and FSA dummies.

	(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
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 9: Average partial effects of incumbency status on being active using coefficients obtained from fixed effects estimator that includes time trends and interaction between time and FSA dummies. Column refers to i while row is for j. Therefore, item (i, j) refers to the effect that j's past incumbency status has on i's decision to be active. Below each chain's name, I indicate their average probability of being active.

	A & W	Burger King	Harvey's	McDonald's	Wendy's
	0.18	0.12	0.13	0.44	0.11
A & W incumbent	0.9436	0.0038	0.0051	0.0101	0.0165
Burger King incumbent	0.0555	0.9707	0.0212	0.0488	0.0001
Harvey's incumbent	-0.0020	0.0128	0.9594	-0.0098	0.0158
McDonald's incumbent	0.0043	0.0116	0.0269	0.8442	0.0234
Wendy's incumbent	0.0521	0.0209	0.0019	0.0136	0.9754

Table 10: Average partial effects of incumbency status on entering a new market using coefficients obtained from fixed effects estimator that includes time trends and interaction between time and FSA dummies. Column refers to i while row is for j. Therefore, item (i, j) refers to the effect that j's past incumbency status has on i's decision to enter. Below each chain's name, I indicate their average probability of entering a market.

	A & W	Burger King	Harvey's	McDonald's	Wendy's
	0.02	0.01	0.01	0.04	0.01
A & W incumbent		0.0005	0.0019	0.0028	0.0057
Burger King incumbent	0.0101		0.0040	0.0171	-0.0009
Harvey's incumbent	0.0001	0.0039		-0.0029	0.0027
McDonald's incumbent	0.0096	0.0032	0.0089		0.0081
Wendy's incumbent	0.0146	0.0070	0.0026	0.0037	

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 active statuses. From Table 8: A & W's decision to enter is positively affected by Burger King's incumbency status; Burger King's decision to enter is positively affected by McDonald's and Wendy's incumbency status; Harvey's decision to enter is positively affected by McDonald's incumbency status; and Wendy's decision to enter is positively affected by A & W and McDonald's incumbency status.

4 Model of entry/exit with learning

4.1 Basic setting

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 (and independently), whether or not to be active in a market m. 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. Let $a_{imt} \in \{0,1\}$ indicate whether chain i is active ($a_{imt} = 1$) or not active ($a_{imt} = 0$) during time t. Choosing not to be active at time t yields a one-shot payoff of zero. Being active in a market yields

$$\Pi_{imt}(a_{imt} = 1) = S^*_{mt}(\theta_{1i} - \sum_{j \neq i} \theta_{2ij} a_{jmt}) - FC_i - (1 - a_{imt-1})EC_i - \varepsilon_{imt}. \tag{4}$$

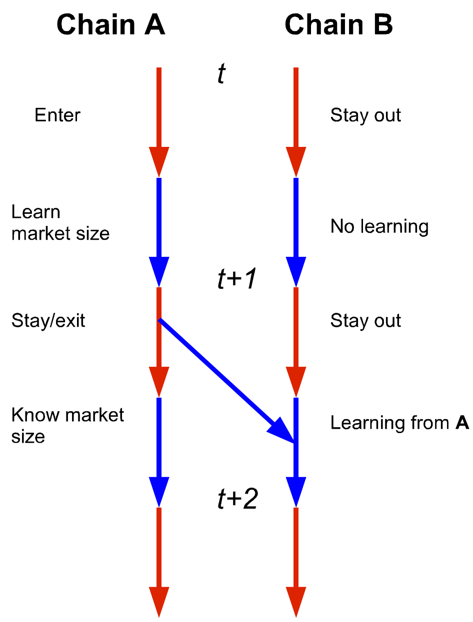
Market size is denoted by $S^*_{mt} = \boldsymbol{Z}_{mt}\boldsymbol{\psi}$, where $\boldsymbol{Z}_{mt}$ is a vector of exogenous market characteristics.[33] The parameter θ_{1i} captures a chain specific fixed effect for revenue; in other words, how effective a chain is at turning potential demand into 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}. As in Seim (2006), each chain receives a privately known and idiosyncratic shock ε_{imt}, which is assumed to be from a type I extreme value distribution. There are also entry and fixed costs, denoted by EC_i and FC_i respectively.

4.2 Information externality

Before describing the information structure in detail, I will illustrate the learning mechanism using a simple example. Suppose that a fast food executive has to decide whether to enter a market that

[33] As in Bresnahan and Reiss (1991), I normalize the coefficient in front of $Population_{mt}$ to be 1, so that variable profit is measured in population units.

Figure 4: Timeline of decisions and information flows.



was recently pitched by a land developer. At his disposal are data on population, income, and the like. Although this information is useful in assessing whether the prospective market has potential, he is not entirely sure about his forecast about market size.[34] While his forecast may be correct, there is also a chance that he underestimates or overestimates the market's potential. If he decides to enter the market, the uncertainty will be resolved; for example, the executive can observe that store's revenue data after entry. Alternatively, he can hold back the investment, and update his prior forecast using past exit/stay decisions of rivals. Because incumbent rivals who decide to stay or exit base their decisions on the true market size, their decisions may in fact be informative to outsiders. This intuition is summarized in Figure 4.

Although the true market size is S^*_{mt}, each chain's expectation of market size is based on the true market size plus some uncertain component,

$$S_{mt} = S^*_{mt} + \omega_m. \tag{5}$$

[34] My assumption that uncertainty is related to demand is largely motivated by Yang's (2010) study, along with conversations with industry executives. As the strategic interaction parameters (θ_{ij}) are determined by the substitutability between products, one would expect fast food chains to be quite certain about how they compare with their rivals.

The term ω_m captures noise in their assessment of market size, which is unknown to the chains initially. They however have common prior beliefs about ω_m, as characterized by the cumulative distribution function $H(\omega_m)$, where

$$H(\omega_m) = \begin{cases} \lambda_0 & \text{if} \quad \omega_m = \sigma \\ \\ 1 - \lambda_0 & \text{if} \quad \omega_m = 0 \end{cases} \tag{6}$$

For simplicity, I assume that the distribution for prior beliefs has a two element support,[35] $\Omega = \{\sigma, 0\}$, and is characterized by the tuple (λ_0, σ). We see that with probability λ_0 the chain has an *incorrect* assessment of market size and with probability $1 - \lambda_0$, a chain's assessment is *correct*. This specification carries two main assumptions: stationarity of ω_m, and common (λ_0, σ) across chains.

These assumptions do come with some caveats. It is possible that over time, prior beliefs naturally become more precise; however, this feature will introduce nonstationarity to the dynamic game, thereby complicating computation, identification, and estimation. The second assumption rules out the possibility that certain firms may have more precise priors than others (i.e., λ_0 close to zero for certain chains). Retail chains, such as McDonald's, are known to have superior real estate research divisions. Alternatively, some chains may be able to learn from their own experiences in similar markets. Relaxing this second assumption is not trivial, as we would have to allow the beliefs regarding the distribution of ω_m to be private information; for example, McDonald's would never share its real estate research with others. Therefore, we would need a rich enough model to allow chains to not only infer the common market uncertainty from an incumbent's exit/stay decision, but also allow them to infer their rival's prior regarding the market. This complication will also obfuscate our notion (and consequently identification) of learning from others, as potential entrants could now learn from their rivals' decisions to not enter particular markets. Although these assumptions make potentially strong abstractions away from the true learning process, the model I have specified is rich enough to capture heterogeneity in learning effects (i.e., which firms care about learning), while at the same time, is simple enough to establish identification.

Each chain's belief about this risk may change over time through own learning and learning from others. Let (λ_{mt}, σ) be chain i's posterior belief at time t, where λ_{mt} is the posterior probability that the assessment is incorrect, and $1 - \lambda_{mt}$ is the posterior probability that the assessment is correct.

[35] The two-point support I use is similar to one of the specifications used by Chernew et al (2008) in their empirical paper on learning and health plans. Note that such a specification of prior beliefs implicitly assumes that the prior distribution for ω is stationary, and that prior beliefs are the same across firms.

A fast food chain will learn the true market size upon entry, which takes place at the beginning of a period t. By learning about the true market size, the incumbent chain i will have an unbiased estimate of market size by the next period $(t+1)$, i.e. $\lambda_{mt+1} = 0$; consequently, its decision to exit/stay at the beginning of $t+1$ is made using this new information. At $t+2$, a rival potential entrant will observe the incumbent's decision at $t+1$, and will try to infer a new set of beliefs based on the observation. Therefore, own learning completely resolves the uncertainty in one period, while learning from others takes two periods to take effect. Consequently, i has no opportunity to learn vicariously as a potential entrant if $a_{jmt-2} = 0$. However, if $a_{jmt-2} = 1$ for at least some j such that, then a potential entrant can update its past beliefs using Baye's rule. First define the set of informed rivals as

$$J^*_{mt} = \{k : a_{kmt-2} = 1\}. \tag{7}$$

Each firm i knows that every firm in the set J^*_{mt} is aware of the true value of ω_m at period $t-1$. Therefore, i knows that every firm in this set has the same value of λ_{mt}, which is either 0 or 1. With this notation in place and using Bayes rule, a potential entrant can then update its beliefs λ_{mt-1} using the following equation:

$$\lambda_{mt} = \frac{\Pr(\boldsymbol{a}_{mt-1}|\lambda = 1)\lambda_{mt-1}}{\Pr(\boldsymbol{a}_{mt-1}|\lambda = 1)\lambda_{mt-1} + \Pr(\boldsymbol{a}_{imt-1}|\lambda = 0)(1-\lambda_{mt-1})}. \tag{8}$$

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

$$\Pr(\boldsymbol{a}_{mt-1}|\lambda = x) = \prod\nolimits_{j\in J^*_{mt}} P_{jm}(\lambda)^{a_{jmt-1}} \cdot (1 - P_{jm}(\lambda))^{(1-a_{jmt-1})} \tag{9}$$

where $P_{jm}(\boldsymbol{\lambda}) = \Pr(a_{jmt} = 1|\lambda_{mt} = \lambda\boldsymbol{)}$. The probability $\Pr(\boldsymbol{a}_{mt-1}|\lambda_{mt} = x)$ captures the information content associated with observed $\boldsymbol{a}_{mt-1}$, which is a vector of actions at period $t-1$ of thse firms that belong to the set J^*_{mt}. Ultimately, whether the rival incumbent has stayed or exited at $t-1$ will affect how the posterior is updated at t.

I now summarize the cases to be considered when updating the posterior beliefs (λ_{mt}, σ):

1. $\{a_{imt-1} = 1\}$ or $\{a_{imt-1} = 0, \lambda_{mt-1} = 0\}$: True market size is known, so $\lambda_{mt} = 0$. The second condition ensures that a chain that learned the true market size will not forget.

2. $\{a_{imt-1} = 0, \lambda_{mt-1} \neq 0, J^*_{mt} = \varnothing\}$: No new information to update posterior, so $\lambda_{mt} = \lambda_0$.

3. $\{a_{imt-1} = 0, \lambda_{mt-1} \neq 0, J^*_{mt} \neq \varnothing\}$: Posterior belief λ_{mt} is updated using the updating equation above.

With this belief updating, all firms that have not entered will hold the same belief, and all firms who have already entered will hold the same belief. Given the posterior beliefs λ_{mt}, chain i's assessment of market size is

$$\begin{aligned} E(S_{mt}) &= \lambda_{mt}(S^*_{mt} + \sigma) + (1 - \lambda_{mt})S^*_{mt} \\ &= S^*_{mt} + \lambda_{mt}\sigma. \end{aligned} \tag{10}$$

Referring back to my illustrative example with chains A and B, I will describe the evolution of beliefs. At time t, chain A has entered and chain B has decided to stay out. When these chains made their decisions at t, they both shared the same beliefs λ_0. Because chain A entered the market at time t, its belief in the subsequent periods will be 0, as its uncertainty has been completely resolved. Although chain B, at time $t+1$, observes chain A's past decision to enter, this past decision has no informative value as the decision was not made when the uncertainty was resolved; therefore, chain B's beliefs do not change: $\lambda_{mt+1} = \lambda_0$. Chain B will update its posterior at time $t+2$, as it observed chain A's decision to stay or exit at time $t+1$ upon resolving its uncertainty. Consequently, $\lambda_{mt+2} = \Pr(a_{mt-1}|\lambda_{mt-1} = 1)\lambda_0/(\Pr(a_{Amt-1}|\lambda_{mt-1} = 1)\lambda_0 + \Pr(a_{mt-1}|\lambda_{mt-1} = 0)(1 - \lambda_0))$.

4.3 Markov Perfect Equilibrium

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

$$\boldsymbol{X}_{mt} = \{\boldsymbol{a}_{mt-2}, \boldsymbol{a}_{mt-1}, \lambda_{mt-1}, \boldsymbol{Z}_{mt}\} \tag{11}$$

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

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

where $V^{\varphi}_i(\boldsymbol{X}_{mt+1}, \varepsilon_{imt+1})$ is the Bellman equation defined as

$$V^{\varphi}_i(\boldsymbol{X}_{mt}, \varepsilon_{imt}) = \max_{a_{imt} \in \{0,1\}} E\left[\Pi^{\varphi}_{imt} + \beta\, V^{\varphi}_i(\boldsymbol{X}_{mt+1}, \varepsilon_{imt+1}) \mid \boldsymbol{X}_{mt}, \varepsilon_{imt}\right]. \tag{13}$$

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

$$P_i(\boldsymbol{X}_{mt}) = \int_i \varphi(\boldsymbol{X}_{mt}, \varepsilon_{imt}) dG_i(\varepsilon_{imt}). \tag{14}$$

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

$$E(\Pi_{imt}^{\varphi} \mid \boldsymbol{X}_{mt}, \varepsilon_{imt}) = \ a_{imt}[\Pi_i^{\mathbf{P}}(\boldsymbol{X}_{mt}) - \varepsilon_{imt}] \tag{15}$$

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

$$\Pi_i^{\mathbf{P}}(\boldsymbol{X}_{mt}) \equiv (S_{mt}^* + \lambda_{imt}\sigma) \left[\theta_{1i} - \sum_{j \neq i} \theta_{2ij} P_j(\boldsymbol{X}_{mt})\right] - FC_i - (1 - a_{imt-1})EC_i. \tag{16}$$

The expectation of the Bellman equation depends on the state vector, transition probability vector[36] $F_i^{\boldsymbol{X},\boldsymbol{P}}(a_{imt}, \boldsymbol{X}_{mt})$, and integrated value functions $\bar{\boldsymbol{V}}_i^{\mathbf{P}}$

$$E(V_i^{\varphi}(a_{imt}, \boldsymbol{X}_{mt+1}, \varepsilon_{imt+1}) | \boldsymbol{X}_{mt},\ \varepsilon_{it}) \equiv F_i^{\boldsymbol{X},\boldsymbol{P}}(a_{imt}, \boldsymbol{X}_{mt})' \bar{\boldsymbol{V}}_i^{\mathbf{P}}. \tag{17}$$

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

$$\bar{V}_i^{\mathbf{P}}(\boldsymbol{X}_{mt+1}) \equiv \int V_i^{\varphi}(\boldsymbol{X}_{mt+1}, \varepsilon_{it+1}) dG(\varepsilon_{it+1}). \tag{18}$$

The best response function for firm i is now defined as

$$\varphi_i(\boldsymbol{X}_{mt}, \varepsilon_{it}) = \begin{cases} 1 & \text{if } \Pi_i^{\mathbf{P}}(\boldsymbol{X}_{mt}) + \beta F_i^{\boldsymbol{X},\boldsymbol{P}}(1, \boldsymbol{X}_{mt})' \bar{\boldsymbol{V}}_i^{\mathbf{P}} \geq \beta F_i^{\boldsymbol{X},\boldsymbol{P}}(0, \boldsymbol{X}_{mt})' \bar{\boldsymbol{V}}_i^{\mathbf{P}} + \varepsilon_{imt} \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Consequently, the best response functions will satisfy

$$P_i(\boldsymbol{X}_{mt}) = G_i\left(\Pi_i^{\mathbf{P}}(\boldsymbol{X}_{mt}) + \beta[F_i^{\boldsymbol{X},\boldsymbol{P}}(1, \boldsymbol{X}_{mt}) - \ F_i^{\boldsymbol{X},\boldsymbol{P}}(0, \boldsymbol{X}_{mt})]' \bar{\boldsymbol{V}}_i^{\mathbf{P}}\right). \tag{20}$$

4.4 Extension: Inclusion of demand externalities

The model presented thus far contains no demand externality. In this section, I describe an extension that will include this feature. An important assumption I make regarding the demand assumption is that it operates through the market size component. Using a similar specification as Toivanen and Waterson (2005), I let

[36] I use the Tauchen (1986) to obtain the transition matrix for the discretized exogenous market characteristics.

$$S^*_{imt} = \boldsymbol{Z}_{mt}\boldsymbol{\psi} + \kappa_i \sum_{j \neq i} a_{jmt-1} \tag{21}$$

where κ_i captures the effect that a demand externality has on chain i. If $\kappa_i > 0$, then the presence of rivals can increase market size;[37] such an effect is possible if consumers benefit from a larger selection of brands or if chains located in close proximity benefit from softened price competition. Unlike the information externality, the rival's past presence alone will have a direct impact on the market size, regardless of whether the rival was active or not at time period $t-2$.

5 Identification of information externalities

5.1 What is the source of uncertainty?

Uncertainty is an important pre-requisite of learning. I offer one possible interpretation about uncertainty under the context of Canada's hamburger fast food industry. According to insight gathered from numerous interviews with executives,[38] the fast food industry faces considerable uncertainty in demand, as they rarely know *ex ante* whether certain consumer segments will be receptive to American style hamburgers. Consumers may vary in taste preferences, as dictated by differences in their cultural background. It is well known that in the 1970s, American style hamburgers were not heavily marketed to markets outside of America (Love, 1995). Also note that Canada has a lot of cultural diversity among its citizens; and as people of similar cultural backgrounds tend to live in the same neighborhoods, much of the uncertainty these managers face will be geographic in nature.

5.2 How is uncertainty resolved?

The main assumption used throughout my model is as follows: once a chain opens an outlet in a particular market, the geographic risk associated with that market will be resolved. This assumption is intuitive. General managers often have access to detailed revenue data from each outlet's sales; consequently, any demand risk they faced prior to store opening should eventually be resolved upon entry. For this assumption to be valid, we need the demand risk to be stationary over time; otherwise, the revenue information they obtained will not shed any light about the uncertainty they face across markets. Manski (2004) makes a similar assumption where outcome distributions exhibit stationarity.

[37] Similar externalities have been found in two-sided markets, where the number of sellers increases the probability of potential sellers to enter (Tucker and Zhang, 2010).

[38] Refer to the Appendix for further details.

Table 11: Example of payoffs from being active for A & W as either uninformed potential entrant or informed incumbent.

		A & W	
		Potential entrant	Incumbent
Burger King	Stay	A	C
	Exit	B	D

Sequential learning requires the observation of past actions. Following Manski (2004), I assume that fast food chains are aware of each market's set of incumbents when they make their entry/exit decisions; a reasonable assumption as fast food outlets are clearly visible to all. Furthermore, outlets are listed in yearly editions of phone directories and retail chain directories once they are open for business.

Incumbent rivals who decide to stay or exit base their decisions on revealed information; consequently, their decisions may be informative to uninformed potential entrants. While potential entrants have an incentive to learn from others, informed incumbents do not. One will not find such an asymmetry between potential entrants and incumbents in their reaction to demand externalities or unobserved heterogeneity.

To some extent, this identification strategy is inspired by past studies that use the following logic: learning from others is more (less) pronounced for cases in which the prior is diffuse (precise). For movie consumption, Moretti (2010) uses whether a movie is a sequel or not; Conley and Udry (2010) use whether farmers are already familiar with a particular production process; Guiso and Schivardi (2007) presume that large firms are more informed than small firms; and for Twitter adoption, Chi and Yang (2010) use whether a politician has already adopted Facebook. All of these examples use a precise prior as a placebo.

5.3 Simple example

I will now illustrate the identification strategy using a simple example. Let us focus on two chains, say, A & W and Burger King. Suppose that we are interested in identifying the effect that Burger King's information externality has on A & W. Note that A & W can either be an uninformed potential entrant, or informed incumbent. Burger King either stayed or exited the market the year before, and this decision would have implications on A & W's payoffs of being active as described in Table 11.

We see from the table that the difference A - B is the difference in payoffs for A & W from entering the market as a potential entrant in light of Burger King's decision to stay as opposed to exit, while the difference C - D is the difference in payoffs for A & W from staying in the market as an incumbent in light of Burger King's decision to stay as opposed to exit. The first difference will capture both learning and non-learning effects, while the second difference will only capture non-learning effects as A & W reacts to informational spillovers only as a potential entrant. Therefore, the double-difference (A - B) - (C - D) will help identify the information externalities. In other words, if A & W is much more likely to enter as a potential entrant, than stay as an incumbent in light of Burger King's decision to stay as opposed to exit, then learning exists.

5.4 Double-difference decomposition

I will now establish the conceptual framework behind the strategy for identifying learning effects. Suppose now that chain i can either be a potential entrant ($a_{imt-1} = 0$) or an incumbent ($a_{jmt-1} = 1$), while one if its rivals, j was an incumbent at time $t-1$ ($a_{jmt-2} = 1$); thus chain i observes j's past decision to stay ($a_{jmt-1} = 1$) or exit ($a_{jmt-1} = 0$). As an incumbent, j's decision at $t-1$ was made without facing uncertainty. The main identification problem is to determine whether i's reaction to j's stay/exit decision at $t-1$ is driven primarily by an information externality. To ensure that i did not already have an opportunity to learn from its or others' past decisions, the market in question was never previously explored by i, j, or other chains prior to $t-2$ ($\{a_{kmt-s} = 0\}_{s>2,\forall k}$).

Define $V_1(a_{imt-1}, a_{jmt-1})$ as the value of being active in a market conditional on (a_{imt-1}, a_{jmt-1}), and $V_0(a_{imt-1}, a_{jmt-1})$ as the value of not being active in a market.[39] There are unobservable components to being active and inactive which I denote as ε_1 and ε_0. Therefore, conditional on its own incumbency status, and the stay/exit decision of its rival, the probability that chain i is active in a market can be defined as:

$$\begin{aligned} P(a_{imt-1}, a_{jmt-1}) &= \Pr(V_1(a_{imt-1}, a_{jmt-1}) + \varepsilon_1 > V_0(a_{imt-1}, a_{jmt-1}) + \varepsilon_0) \\ &= \Pr(\tilde{\varepsilon} < \tilde{V}(a_{imt-1}, a_{jmt-1})) \\ &= F_{\tilde{\varepsilon}}(\tilde{V}(a_{imt-1}, a_{jmt-1})) \end{aligned} \tag{22}$$

where $\tilde{\varepsilon} = \varepsilon_0 - \varepsilon_1$, $\tilde{V}(a_{imt-1}, a_{jmt-1}) = V_1(a_{imt-1}, a_{jmt-1}) - V_0(a_{imt-1}, a_{jmt-1})$, and $F_{\tilde{\varepsilon}}(\cdot)$ is the cumulative distribution function for the unobservable $\tilde{\varepsilon}$. Note that given a parametric assumption re-

[39] Market characteristics also enter the value of being active. For notational simplicity, I take as assumed that $V_1(\cdot)$ also depends on a vector of market characteristics $\boldsymbol{Z}_{mt}$.

garding $F_{\tilde{\varepsilon}}(\cdot)$, and data on entry/exit decisions, $\tilde{V}(a_{imt-1}, a_{jmt-1})$ is identified semi-parametrically:

$$\tilde{V}(a_{imt-1}, a_{jmt-1}) = F_{\tilde{\varepsilon}}^{-1}(P(a_{imt-1}, a_{jmt-1})) \tag{23}$$

If we assume that $\tilde{\varepsilon}$ has a $U[0,1]$ distribution, then $\tilde{V}(a_{imt-1}, a_{jmt-1}) = P(a_{imt-1}, a_{jmt-1})$. Alternatively, we can let $\tilde{\varepsilon}$ be a logit error, in which case $\tilde{V}(a_{imt-1}, a_{jmt-1}) = \log(P(a_{imt-1}, a_{jmt-1})/[1-P(a_{imt-1}, a_{jmt-1})])$. With this notation in place, I place some additional structure on the identifiable $\tilde{V}(a_{imt-1}, a_{jmt-1})$. Here, I write $\tilde{V}(a_{imt-1}, a_{jmt-1})$ as

$$\tilde{V}(a_{imt-1}, a_{jmt-1}) = E(\tilde{R}|a_{imt-1}, a_{jmt-1}). \tag{24}$$

where $E(\tilde{R}|a_{imt-1}, a_{jmt-1})$ is chain i's expectation about the net benefit of being active conditional on its own incumbency status, as well as its rival's past decision to stay or exit.

The first difference in my double-difference decomposition is defined by:

$$\tilde{V}(0,1) - \tilde{V}(0,0) = E(\tilde{R}|0,1) - E(\tilde{R}|0,0). \tag{25}$$

This object captures a **potential entrant's** reaction to its rival's past decision to stay or exit. Note that this difference can potentially be non-zero for two reasons. First, a rival's past presence could be indicative of either unobserved heterogeneity and/or demand externalities. Secondly, a rival's past presence could affect the potential entrant's information set, as it still faces uncertainty. In other words, this first difference will capture both learning and non-learning effects induced by j's decision to stay as opposed to exit. The second difference in my double-difference decomposition is defined by:

$$\tilde{V}(1,1) - \tilde{V}(1,0) = E(\tilde{R}|1,1) - E(\tilde{R}|1,0). \tag{26}$$

This object captures an **incumbent's** reaction to its rival's past decision to stay or exit. Now, this difference can potentially be non-zero for only one reason, namely that a rival's past presence is representative of either unobserved heterogeneity and/or demand externalities. Unlike a potential entrant, an incumbent's net benefit of being active will only be affected by non-learning factors as it no longer faces uncertainty. Thus, this second difference captures only non-learning effects induced by j's decision to stay as opposed to exit. Subtracting the second difference from the first difference yields:

$$\overbrace{[\tilde{V}(0,1) - \tilde{V}(0,0)]}^{\text{Potential entrant}} - \overbrace{[\tilde{V}(1,1) - \tilde{V}(1,0)]}^{\text{Incumbent}} \tag{27}$$
$$= \underbrace{\{[E(\tilde{R}|0,1) - E(\tilde{R}|0,0)] - [E(\tilde{R}|1,1) - E(\tilde{R}|1,0)]\}}_{\text{Learning effect}}$$

Therefore, taking the double-difference will allow us to net out the non-learning effects that affect both potential entrants and incumbents alike, and thus, isolate the learning effect. Note that the double-difference will yield a lower bound if we allow for endogenous sunk costs (Sutton, 1991) and/or learning-by-doing. The argument for this assertion is that such features will make the competition effect associated with a rival's past presence more intense for potential entrants relative to incumbents. As the fast food industry is advertising intensive, the past presence of a rival can raise the cost of entry; this effect works in the opposite direction as learning. With learning-by-doing, an incumbent chain is more likely to compete with rivals as it will have moved down its cost curve with experience, thereby allowing it to price competitively. If we fail to account for these two cases, $\tilde{V}(0,1) - \tilde{V}(0,0)$ will be biased downwards.

5.5 Implementation

I obtain the values of $P(1,1)$, $P(1,0)$, $P(0,1)$, and $P(0,0)$ for each chain-to-chain interaction using either a simple probability frequency or logit estimator that conditions on a chain's incumbency status, its rival's incumbency status, and the observed market characteristics. For the frequency estimator, I categorize all of the continuous market characteristics into 0.25, 0.5, and 0.75 quartiles so that each vector element in the support $\boldsymbol{Z} = \{\boldsymbol{Z}_{mt}\}_{\forall m,t}$ has a non-trivial probability of being observed.

I will now go over the number of observations that satisfy each of the four main conditions, namely the size of subsamples that satisfy $a_{imt-1} = a_i$, $a_{jmt-1} = a_j$, $a_{jmt-2} = 1$, and $\{a_{kmt-s} = 0\}_{s>2,\forall k}$ for all $(4 \times 5) \times 2^2 = 80$ possible combinations of chain-chain pairings. In general, we see that the data provides the largest subsamples for identifying $P(0,1)$. In contrast, the subsamples used to identify $P(1,1)$, $P(1,0)$ and $P(0,0)$ are smaller.[40] There are disproportionately more cases in which an incumbent stays, than exits; this feature explains why there are so many more observations in the subsample for $P(0,1)$ in comparison to subsamples for $P(1,0)$ and $P(0,0)$. The condition that no other chain had been active in the market makes the subsample for $P(1,1)$ much

[40] Subsample sizes used to estimate $P(1,1)$, $P(1,0)$, $P(0,1)$, and $P(0,0)$ for each chain-to-chain interaction are on average around 40, 60, 200, and 15.

Table 12: The role of information externalities obtained with the assumption that $\tilde{\varepsilon}$ is uniformly distributed. Here, column pertains to chain i, and row pertains to i's rival j. Note that the information externality associated with McDonald's past stay/exit decisions is not identified as their exit is virtually negligible. Below each chain's name, I indicate their average probability of being active.

	A & W	Burger King	Harvey's	McDonald's	Wendy's
	0.18	0.12	0.13	0.44	0.11
A & W		0.02	-0.14	0.02	0.01
Burger King	0.03		-0.18	0.01	0.01
Harvey's	0.02	0.02		-0.01	0.02
McDonald's	N/A	N/A	N/A		N/A
Wendy's	0.03	0.03	0.03	0.01	

smaller than the subsample for $P(0,1)$. If the decision making chain was active during the last period, it is very likely it was also active in periods further back in time. Therefore, such a chain would be removed from consideration when looking at the impact its rival's decision has on the decision to stay/exit.

Also note that, there is virtually no exit for McDonald's, which precludes me from identifying $\hat{P}(a_i, a_{MCD})$ for all $i \neq MCD$. Therefore, the empirical strategy I have proposed does not allow me to identify the information externalities coming from McDonald's past stay/exit decisions. However, its decision to expand or not can be used to identify learning.[41]

5.6 Evidence of information externalities

My differencing approach reveals that information externalities can explain the observed clustering behavior we see in Canada's fast food industry. For my first set of estimates, I assume that $\tilde{\varepsilon}$ has a uniform distribution. Table 12 shows that for many chain-to-chain interactions, the information externality has a positive impact on entry.[42] In many cases, the information externality increases the probability of being active by 2 percentage points. To evaluate the magnitude of this effect, note that a 1% increase in the population leads to a 3 percentage point increase in the probability of being active and that the unconditional probability of being active ranges from 10 to 44 percentage points; therefore, this externality can have a real impact on most of the chains.

[41] I exploit this data variation in a modified learning statistic proposed and implemented in the Appendix.

[42] The double-difference decomposition I propose does not easily yield standard errors. However, in the Appendix, I discuss an alternative approach that implements a flexible logit with random fixed effects. The results are qualitatively the same, and in some cases, statistically significant at 10%, 5% or 1% levels. In particular, the results for A & W, Burger King, and Harvey's exhibit the least noise.

Table 13: The role of demand externalities, competition effects and unobserved heterogeneity with the assumption that $\tilde{\varepsilon}$ is uniformly distributed. Here, column pertains to chain i, and row pertains to i's rival j. Note that the demand externalities, competition effects and unobserved heterogeneity associated with McDonald's past stay/exit decisions is not identified as their exit is virtually negligible.

	A & W 0.18	Burger King 0.12	Harvey's 0.13	McDonald's 0.44	Wendy's 0.11
A & W		0.00	0.15	0.00	0.00
Burger King	0.02		0.20	0.00	0.01
Harvey's	0.01	0.01		0.02	0.00
McDonald's	N/A	N/A	N/A		N/A
Wendy's	0.01	0.00	0.00	0.00	

Heterogeneity in learning effects may be consistent with a number of explanations. One possibility is that some chains spend more on real estate research than others, which ultimately reflects varying preferences towards information (i.e., some chains value a reduction of uncertainty more than others). Alternatively, managers may vary in skill across chains.[43] This variance in skill may imply that some managers have in depth knowledge about the markets they are responsible for, while others have little knowledge. Those with little knowledge may have a stronger incentive to take advantage of information externalities associated with the entry/exit decisions of more experienced managers.

Table 13 shows that non-learning effects have relatively smaller impact on chains, with the exception of Harvey's. In fact, these non-learning effects are in fact quite large for Harvey's. The large effects are generated by the observation that Harvey's is much more likely to stay in a market that is occupied by a rival, than in a market that is not occupied by a rival.[44] There are two possible explanations for this result. First, Harvey's may be less efficient at advertising and marketing, and thus, face prohibitive entry barriers in light of rival presence. This explanation is consistent with my earlier discussion about biases associated with asymmetric competition effects among potential entrants and incumbents. Alternatively, Harvey's may simply be adopting a non-discretionary approach towards choosing locations. If Harvey's enters markets that are profitable, as well as markets that are not profitable, then it should only survive in profitable (and inherently popular) markets.

[43]The relationship between manager skill and entry strategies has been well established by Goldfarb and Xiao (2011) in their study about the US telephone industry.

[44]The finding that Harvey's is less likely to exit under the presence of rival incumbents is reassuring in the sense that the exit of Harvey's is not simply the process of "making room" for American fast food chains.

Table 14: The role of information externalities with the assumption that $\tilde{\varepsilon}$ is a logit error. Here, column pertains to chain i, and row pertains to i's rival j. Note that the information externalities associated with McDonald's past stay/exit decisions is not identified as their exit is virtually negligible.

	A & W	Burger King	Harvey's	McDonald's	Wendy's
A & W		0.33	-0.19	-0.15	0.28
Burger King	0.25		0.14	-0.13	0.35
Harvey's	0.14	0.41		0.11	0.15
McDonald's	N/A	N/A	N/A	N/A	N/A
Wendy's	0.02	0.05	-0.06	0.15	

Table 15: The role of demand externalities, competition effects and unobserved heterogeneity with the assumption that $\tilde{\varepsilon}$ is a logit error. Here, column pertains to chain i, and row pertains to i's rival j. Note that the information externality associated with McDonald's past stay/exit decisions is not identified as their exit is virtually negligible.

	A & W	Burger King	Harvey's	McDonald's	Wendy's
A & W		-0.32	-0.07	0.16	0.64
Burger King	0.02		0.27	-0.13	0.98
Harvey's	-0.77	0.41		0.43	0.1
McDonald's	N/A	N/A	N/A		N/A
Wendy's	1.26	-0.25	0.61	0.15	

Similar results hold if I instead assume that $\tilde{\varepsilon}$ is a logit error, as I still find evidence of learning for a number of firms (Table 14). In general, the learning spillovers in most cases outweighs the non-learning spillovers. Except now, the learning effect is less pronounced for McDonald's. As before, I still find that non-learning effects outweigh learning effects for Harvey's (Table 15).

6 Estimation of structural model

6.1 Identification of structural model

The model I have presented is nested within the class of stationary dynamic games with private information that is additively separable. Furthermore, the model incorporates an exclusion restriction, in which a chain's incumbency status directly enters it's own one-shot payoff via the entry cost, and only affects the rival chains indirectly through the learning mechanism. Taken together, these assumptions along with rich variation in entry/exit decisions over a long time period ensure that the proposed model is identified.[45]

As my data does not contain revenue, variable profits are identified using time-varying market characteristics,[46] as in Bresnahan and Reiss (1991). These market characteristics are the same ones used in my earlier reduced form analysis. The evolution of population, income and property value helps me identify the transition probabilities of exogenous characteristics.

Although I have established a clear and comprehensive identification strategy for the overall learning effect, the identification argument for the specific parametrization of uncertainty in my model is a bit more subtle. That said, I will briefly outline the type of variation in data needed in order to separately identify (λ_0, σ). Note that the posterior beliefs can be written as

$$\lambda_{mt} = f^{\mathbf{P}}(\boldsymbol{X}_{mt}) + g^{\mathbf{P}}(\boldsymbol{X}_{mt}) \cdot \lambda_0 \tag{28}$$

where $f^{\mathbf{P}}(\boldsymbol{X}_{mt})$ and $g^{\mathbf{P}}(\boldsymbol{X}_{mt})$ are functions of the state variables and conditional choice probabilities. Consequently, the assessment of market size is

$$E(S_{mt}) = S^*_{mt} + f^{\mathbf{P}}(\boldsymbol{X}_{mt}) \cdot \sigma + g^{\mathbf{P}}(\boldsymbol{X}_{mt}) \cdot \lambda_0 \cdot \sigma. \tag{29}$$

The parameter σ can be identified provided that $f^{\mathbf{P}}(\boldsymbol{X}_{mt})$ and $g^{\mathbf{P}}(\boldsymbol{X}_{mt})$ are not collinear with S^*_{mt}. However, this parameter needs to be separately identified from λ_0. Therefore, one can separate these two parameters if $f^{\mathbf{P}}(\boldsymbol{X}_{mt})$ and $g^{\mathbf{P}}(\boldsymbol{X}_{mt})$ are not collinear. Intuitively, we see that the degree

[45] Refer to Blevins (2010) for a detailed sumary of identification conditions.

[46] Refer to Campbell and Hopenhayn (2005) for analysis showing population as a key driver of market structure.

of uncertainty (σ) appears in a chain's expected payoff provided that its uncertainty has not yet been resolved, and has posterior beliefs $\lambda_{mt} \neq 0$. Identifying σ demands less out of the data, as we only need variation among those that are not incumbents. The parameter λ_0 appears in the payoff only when a potential entrant has not yet updated its beliefs about the market size; that is, I need enough observations for which potential entrants make decisions whether to enter virgin markets. Therefore, λ_0 will most likely appear in the expected payoff during the early years in which potential entrants are naive. This observation reiterates the importance of having data that dates back to an industry's initial stages.

In the model that contains both information and demand externalities, separate identification of the parameters associated with uncertainty and κ_i require even richer variation in the data. To guide my identification argument, I write out the expected market size as

$$E(S_{mt}) = \boldsymbol{Z}_{mt}\boldsymbol{\psi} + \kappa_i \sum_{j \neq i} a_{jmt-1} + f^{\mathbf{P}}(\boldsymbol{X}_{mt}) \cdot \sigma + g^{\mathbf{P}}(\boldsymbol{X}_{mt}) \cdot \lambda_0 \cdot \sigma. \tag{30}$$

In particular, we need $\sum_{j \neq i} a_{jmt-1}$, $f^{\mathbf{P}}(\boldsymbol{X}_{mt})$ and $g^{\mathbf{P}}(\boldsymbol{X}_{mt})$ to not be collinear. There are two features that help with identification. First, variation in $f^{\mathbf{P}}(\boldsymbol{X}_{mt})$ and $g^{\mathbf{P}}(\boldsymbol{X}_{mt})$ depends heavily on how differently potential entrants and incumbents react to a rival's decision to stay or exit, while variation in $\sum_{j \neq i} a_{jmt-1}$ is simply determined by each rival's incumbency status. Second, the Bayesian learning process distinguishes $f^{\mathbf{P}}(\boldsymbol{X}_{mt})$ and $g^{\mathbf{P}}(\boldsymbol{X}_{mt})$ from $\sum_{j \neq i} a_{jmt-1}$ via functional form.

6.2 Modified NPL procedure

The parameters in my model are $\boldsymbol{\chi} = \{FC_i, EC_i, \theta_{1i}, \theta_{2ij}, \boldsymbol{\psi}, \sigma\}_{\forall i}$ and λ_0. Before estimation, I first map the structural parameters in the one-shot payoff into reduced form parameters, $\boldsymbol{\pi} = h(\boldsymbol{\chi})$. This mapping amounts to expanding out the terms in the expected profits. Therefore, conditional on $\boldsymbol{X}_{mt}$ and $\boldsymbol{\Upsilon} = \{\boldsymbol{\pi},\lambda_0\}$, the best response probability function $G_i(\cdot)$ is used to construct the pseudo-likelihood equation,

$$Q(\boldsymbol{\Upsilon}, \boldsymbol{P}) = \sum_{i,m,t} \{a_{imt} \log G_i(\boldsymbol{P}_{-i}(\boldsymbol{X}_{mt}), \boldsymbol{X}_{mt}|\boldsymbol{\Upsilon}) + (1 - a_{imt}) \log[1 - G_i(\boldsymbol{P}_{-i}(\boldsymbol{X}_{mt}), \boldsymbol{X}_{mt}|\boldsymbol{\Upsilon})]\}. \tag{31}$$

This pseudo-likelihood is highly nonlinear in the prior probability λ_0. Therefore, I consider an algorithm that essentially concentrates out $\boldsymbol{\pi}$, and then searches for λ_0 over a grid space. The algorithm is an extension of the Nested Pseudo Likelihood method originally designed by Aguirregabiria and Mira (2007).[47] 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 computationally tractable. Moreover, the NPL estimates are more efficient than alternative two-step methods.[48] The modified NPL algorithm can be described as follows:

1. Generate a grid of possible values for $\lambda_0^{(g)} \in [0, 0.01, 0.02, ..., 0.99, 1]$.
2. Estimate non-parametrically the initial CCP vector $\hat{\boldsymbol{P}}^{0,(g)}$.
3. Given $\boldsymbol{X}_{mt}$, $\hat{\boldsymbol{P}}^{0,(g)}$, and $\lambda_0^{(g)}$, generate a sequence of posterior beliefs for each firm and market $\{\hat{\lambda}_{mt}^{0,(g)}\}_{\forall m,t}$.
4. Given $\boldsymbol{X}_{mt}$, $\hat{\boldsymbol{P}}^{0,(g)}$, $\lambda_0^{(g)}$, and $\{\hat{\lambda}_{mt}^{0,(g)}\}_{\forall m,t}$, find
$$\hat{\boldsymbol{\pi}}^{0,(g)} = \arg\max_{\boldsymbol{\chi}} Q(\{\boldsymbol{\chi}, \lambda_0^{(g)}\}, \hat{\boldsymbol{P}}^{0,(g)} | \boldsymbol{X}_{mt}, \{\hat{\lambda}_{imt}^{0,(g)}\}_{\forall i,m,t}). \tag{32}$$
5. Update $\hat{\boldsymbol{P}}^{0,(g)}$ using $\hat{\boldsymbol{P}}^{1,(g)} = \{G_i(\hat{\boldsymbol{P}}_{-i}^{0,(g)}(\boldsymbol{X}_{mt}), \boldsymbol{X}_{mt} | \boldsymbol{X}_{mt}, \hat{\boldsymbol{\pi}}^{0,(g)}, \lambda_0^{(g)}, \{\hat{\lambda}_{mt}^{0,(g)}\}_{\forall m,t})\}_{\forall m,t}$.
6. Repeat steps 2 to 5 until $\left\| \hat{\boldsymbol{P}}^{k+1,(g)} - \hat{\boldsymbol{P}}^{k,(g)} \right\|$ is close to zero,[49] where k is equal to the number of iterations. Once convergence is reach, we have $\hat{\boldsymbol{P}}^{NPL,(g)}$ and $\hat{\boldsymbol{\pi}}^{NPL,(g)}$.
7. Obtain the structural parameters $\hat{\boldsymbol{\chi}}^{NPL,(g)}$ using the minimum distance estimator:
$$\hat{\boldsymbol{\chi}}^{NPL,(g)} = \arg\min_{\boldsymbol{\chi}} (\hat{\boldsymbol{\pi}}^{NPL,(g)} - h(\boldsymbol{\chi}))' Avar(\hat{\boldsymbol{\pi}}^{NPL,(g)})^{-1} (\hat{\boldsymbol{\pi}}^{NPL,(g)} - h(\boldsymbol{\chi})). \tag{33}$$
8. Do steps 2 to 7 for each possible value for $\lambda_0^{(g)}$, and then choose $\lambda_0^{(g)}$ and $\hat{\boldsymbol{\chi}}^{NPL,(g)}$ that has the highest pseudo-likelihood function.

When generating the sequence of posterior beliefs in Step 3, I face a complication. In the model, Bayesian updating uses each incumbent rival's CCP conditional on either $\lambda_{mt-1} = 0$ or $\lambda_{mt-1} = 1$. However, observing an incumbent rival that was already active at $t-2$ with incorrect beliefs ($\lambda_{mt-1} = 1$) is purely counterfactual. I address this issue by conditioning the rival choice probabilities on the observable data and whether the rival is a potential entrant or incumbent. An uninformed rival's decision to enter or stay out can give us $P_{jmt-1}(\lambda_{mt-1} = 1)$, and an informed rival's decision to stay or exit can give us $P_{jmt-1}(\lambda_{mt-1} = 0)$. Ultimately, variation in market size can help identify these two measures.

[47] This estimation approach has been used in a variety of applications. Some examples include Aguirregabiria (2004), Aguirregabiria and Ho (2009), Ellickson and Misra (2008), Han and Hong (2008), Lenzo (2010), Magesan (2010), Sweeting and Roberts (2010), Walrath (2008), and Yang (2010).

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

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

7 Results

7.1 Summary of structural estimates

My structural estimates[50] are summarized in Table 16. The coefficients for the market characteristics are of similar signs as those in my reduced form analysis. As in the previous section, retail activity and the proportion of those working in the same FSA have positive effects on the expected payoff, while income, population and property value have negative effects. There is also some heterogeneity in terms of each chain's cost structure. For instance, A & W and Wendy's have the largest fixed costs, while McDonald's has the highest entry cost. McDonald's high entry cost could be driven by a multitude of reasons. First, part of this cost 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. Similar to Toivanen and Waterson (2005) and Vitorino (2008), my estimates for strategic interaction suggest a potential for complementarity between chains, as the presence of a rival increases the expected payoff. Therefore, demand externalities may play a role in the overall spillover effect.

Most importantly, I find that chains face uncertainty. There is a significant probability that their *ex ante* assessments are wrong; and the extent to which they are wrong is also large. A positive sign for the degree of uncertainty (σ) also suggests that *ex ante*, chains are overly optimistic. Therefore, the benefit of learning would be to avoid seemingly profitable markets that are in fact, not profitable. As uncertainty is a prerequisite for learning, these results provide indirect evidence of learning. This finding in itself is a contribution to the study of spillover effects in retail, as the status quo has largely produced reduced form estimates of these effects.

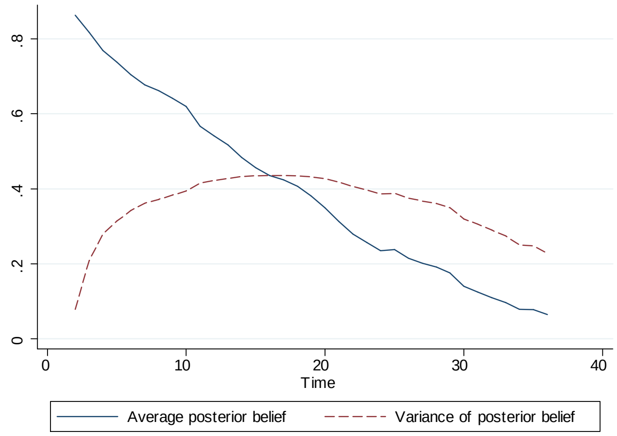
Figure 5 illustrates that the posterior beliefs converge towards $\hat{\lambda}_{mt} = 0$ over time. Potential entrants are learning about the markets vicariously through the observed exit/stay decisions of their rivals. With the inferred posterior beliefs, I can look at how they change over time. What I am interested in is how the variation of these beliefs change over time. The same figure shows that variation in $\hat{\lambda}_{mt}$ for potential entrants initially increases, and then falls over time. In the very beginning, chains beginning with the same common prior across all markets they are uncertain about. As they learn about markets through learning from others, the beliefs across markets begin to diverge. There will be few markets for which $\hat{\lambda}_{mt} > 0$ in the latter years because of learning.

[50] To initiate the portion of my algorithm that uses the NPL, I not only consider a nonparametric estimate of the equilibrium probabilities, but also random draws of these probabilities. By initializing the NPL at different starting points, I can check that the estimation procedure yields unique NPL fixed points. I find that my estimates are the same regardless of the initialization. In general, less than 10 NPL iterations are needed for convergence; so much of the computational cost was associated with the fact that I have to do a separate grid search for λ_0.

Table 16: Structural estimation of entry model with information externalities. Note that the discount factor is calibrated at $\beta = 0.95$. Significance at 5% level denoted by *, and significance at 1% level denoted by **.

Variable	Estimate	Variable	Estimate
Market characteristics		Strategic interaction	
Proportion in same FSA	5.1851** (0.0042)	A & W vs Burger King	-0.0370** (0.0006)
Income	-2.6288** (0.0055)	A & W vs Harvey's	-0.0045 (0.0221)
Total sales	1.6260** (0.0055)	A & W vs McDonald's	0.0075 (0.0055)
Total locations	0.4853** (0.0036)	A & W vs Wendy's	-0.0340** (0.0086)
Population density	-0.2855** (0.0229)	Burger King vs A & W	-0.0238** (0.0010)
Property value	-0.3827** (0.0031)	Burger King vs Harvey's	-0.0232* (0.0116)
Fixed costs		Burger King vs McDonald's	-0.0342* (0.0134)
A & W	1.1760** (0.0149)	Burger King vs Wendy's	-0.0240* (0.0143)
Burger King	0.9602** (0.0088)	Harvey's vs A & W	0.0235 (0.0185)
Harvey's	0.7942** (0.0091)	Harvey's vs Burger King	-0.0236** (0.0072)
McDonald's	-1.1391** (0.0049)	Harvey's vs McDonald's	-0.0237 (0.0122)
Wendy's	1.4675** (0.0120)	Harvey's vs Wendy's	-0.0194** (0.0025)
Entry costs		McDonald's vs A & W	-0.0088 (0.0168)
A & W	4.5577** (0.0371)	McDonald's vs Burger King	-0.0112 (0.0139)
Burger King	4.9283** (0.0009)	McDonald's vs Harvey's	0.0148** (0.0062)
Harvey's	4.4807** (0.0002)	McDonald's vs Wendy's	0.0165 (0.0100)
McDonald's	5.9324** (0.0056)	Wendy's vs A & W	-0.0322* (0.0154)
Wendy's	5.3298** (0.0054)	Wendy's vs Burger King	-0.0138 (0.0203)
Uncertainty		Wendy's vs Harvey's	-0.0350** (0.0077)
σ	0.3974** (0.0021)	Wendy's vs McDonald's	-0.0530* (0.0231)
λ_0	0.89** (0.0010)	θ_{1AW}	0.1171** (0.0054)
		θ_{1BK}	0.0742** (0.0286)
		θ_{1HARV}	0.0562** (0.0145)
LL = -7572.9		θ_{1MCD}	0.1269** (0.0171)
N = 20930		θ_{1WEND}	0.1135** (0.0239)

Figure 5: Evolution of estimated posterior beliefs ($\hat{\lambda}_{mt}$) over time.



This result is interesting, as it confirms the importance of having data on industry dynamics from the early to mid years.

The structural estimates for the model that includes both information and demand externalities appear in Table 17. Even when demand externalities are explicitly included to the model, the parameters associated with uncertainty experience little change. The estimates associated with the demand externalities are worth noting. Demand externalities appear to affect each firm differently. In particular, the demand externality has the largest impact on Wendy's, and smallest impact on McDonald's.

7.2 Implications of uncertainty

The estimated structural model provides us an opportunity to learn more about the role of uncertainty in two aspects: market structure and profitability. To investigate the impact of uncertainty on market outcomes, I compare the market structure or profitability associated with the counterfactual scenario with no uncertainty and the actual scenario with uncertainty. The counterfactual scenario is implemented by setting $\sigma = 0$ so as to calculate counterfactual probabilities using the method proposed by Aguirregabiria (2009). These counterfactual probabilities are then used to generate a new sequence of entry/exit decisions for each market under the scenario of no uncertainty.

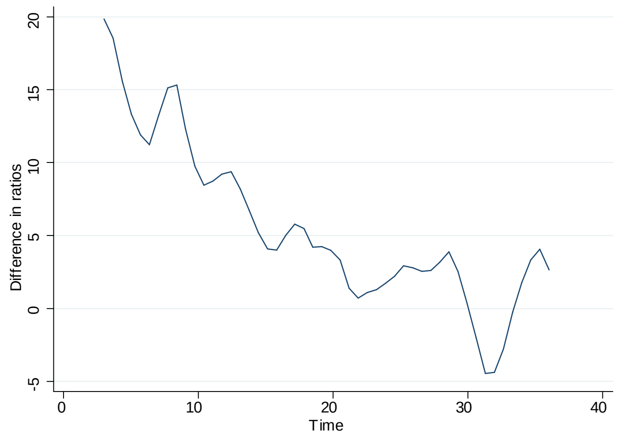
7.2.1 On market structure

The objective of this analysis is to establish a link between uncertainty, learning, and market power. Canada's fast food industry has become increasingly competitive over time (Figure 2), and

Table 17: Structural estimation of entry model with information and demand externalities. Note that the discount factor is calibrated at $\beta = 0.95$. Significance at 5% level denoted by *, and significance at 1% level denoted by **.

Variable	Estimate	Variable	Estimate
Market characteristics		Strategic interaction	
Proportion in same FSA	5.3668** (0.0100)	A & W vs Burger King	-0.0185 (0.0112)
Income	-1.2428** (0.0177)	A & W vs Harvey's	0.0014 (0.0082)
Total sales	1.4977** (0.0029)	A & W vs McDonald's	0.0087** (0.0013)
Total locations	0.9987** (0.0157)	A & W vs Wendy's	-0.0168** (0.0018)
Population density	-0.2806** (0.0300)	Burger King vs A & W	-0.0095* (0.0047)
Property value	-0.9005** (0.0050)	Burger King vs Harvey's	-0.0097 (0.0133)
Fixed costs		Burger King vs McDonald's	-0.0140 (0.0293)
A & W	2.0696** (0.0056)	Burger King vs Wendy's	-0.0119 (0.0115)
Burger King	1.3514** (0.0010)	Harvey's vs A & W	0.0163 (0.0136)
Harvey's	1.1402** (0.0018)	Harvey's vs Burger King	-0.0124 (0.0149)
McDonald's	-0.7722** (0.0058)	Harvey's vs McDonald's	-0.0115 (0.0129)
Wendy's	2.1772** (0.0011)	Harvey's vs Wendy's	-0.0110 (0.0382)
Entry costs		McDonald's vs A & W	-0.0059 (0.0426)
A & W	4.9976** (0.0261)	McDonald's vs Burger King	-0.0057* (0.0013)
Burger King	5.1038** (0.0147)	McDonald's vs Harvey's	0.0096* (0.0033)
Harvey's	4.6499** (0.0056)	McDonald's vs Wendy's	0.0093* (0.0045)
McDonald's	6.2685** (0.0039)	Wendy's vs A & W	-0.0078* (0.0028)
Wendy's	5.7113** (0.0059)	Wendy's vs Burger King	0.0007 (0.0163)
Uncertainty		Wendy's vs Harvey's	-0.0089 (0.0071)
σ	0.6532** (0.0033)	Wendy's vs McDonald's	-0.0184** (0.0005)
λ_0	0.97** (0.0021)	θ_{1AW}	0.1297** (0.0025)
Demand externality		θ_{1BK}	0.0711** (0.0076)
A & W's benefit	0.5189** (0.0052)	θ_{1HARV}	0.0570** (0.0094)
Burger King's benefit	1.1230** (0.0041)	θ_{1MCD}	0.1067** (0.0262)
Harvey's benefit	0.5118** (0.0032)	θ_{1WEND}	0.1177** (0.0004)
McDonald's benefit	-0.0196** (0.0010)	LL = -7578.7	
Wendy's benefit	1.9699** (0.0021)	N = 20930	

Figure 6: The ratio between the total number of markets with one firm and the total number of markets with more than one firm.



this analysis will inform us to what extent uncertainty and learning contribute to this phenomenon. I compare the actual market structure ($\sigma \neq 0$) with the counterfactual market structure ($\sigma = 0$) over time. To make this comparison, I compare the competitiveness for each regime using the following ratio:

$$\frac{\#\text{ Monopoly Markets}}{\#\text{ Oligopoly Markets}}. \tag{34}$$

Figure 6 plots the difference between the ratio associated with an industry with uncertainty, and the ratio associated with an industry without uncertainty. During the first 20 years, the ratio is larger under the actual scenario with uncertainty than the counterfactual scenario without uncertainty; in other words, an industry with uncertainty is less competitive than an industry with certainty initially. However, the ratio associated with uncertainty converges to the ratio associated without uncertainty during the latter years. This pattern is consistent with the idea of learning from the actions of competitors: uncertainty may be an entry barrier initially, but the observation of successful incumbents encourage new entry in the latter years, which ultimately leads to greater increases in the number of competitors for each market (with active incumbents) and convergence towards a competitive outcome.

7.2.2 On entry and exit

Firms may be motivated to learn from the past decisions of rivals if doing so allows them to avoid markets that are most likely to be unprofitable. I explore this motivation in greater detail, and investigate whether entry and exit are more prevalent when uncertainty is present. If firms face

Table 18: Measure of excessive entry/exit under uncertainty.

	1970 - 1990	1990 - 2005
Entry	3.7%	2.8%
Exit	7%	6.5%

uncertainty, they may be more prone to making incorrect assessments about market size, ultimately leading to excessive entry into *ex post* sub-optimal markets. Upon revelation of the true market size, they may wish to exit, especially when their forecasts were overly optimistic. I make the comparison between exit rates with and without uncertainty by using the following metric:

$$\frac{E(N^{Uncertainty}) - E(N^{Certainty})}{E(N^{Certainty})}. \tag{35}$$

Here, $E(N^{Uncertainty})$ is the average number of chains that enter/exit when they face uncertainty, and $E(N^{Certainty})$ is the average number of chains that enter/exit when they do not face uncertainty. Therefore, this quantity captures how much more entry/exit there may be in light of uncertainty. I calculate this quantity separately for the initial years (1970 to 1990), and the latter years (1990 to 2005).

Table 18 shows us that when an industry exhibits uncertainty, chains enter markets excessively and are much more likely to exit. Furthermore, the disproportionate amount of entry/exit under uncertainty is more pronounced during the early years (1970 to 1990). This pattern reflects the idea that uncertainty plays a greater role initially, as firms have fewer opportunities to learn from others before making their entry decisions; thus, their decisions to enter are more likely to be based on incorrect assessments about market size, ultimately resulting in *ex post* failure. To summarize, there are many instances in which chains enter markets that they would have avoided *ex post* had they known the true market size, and such cases are less prevalent over time as firms are able to make inferences about the uncertain market size via learning.

7.2.3 On profitability

As I have shown in the Appendix, the statistic for learning is large only to those that face uncertainty. Given this relationship between the degree of uncertainty and my measure of learning, a natural question would be whether firms value a reduction in uncertainty; that is, does *ex ante* uncertainty act like an entry barrier? Furthermore, is there a close correspondence between the value they place towards certainty and their reaction to information externalities?

Figure 7: The value associated with uncertainty reduction over time.

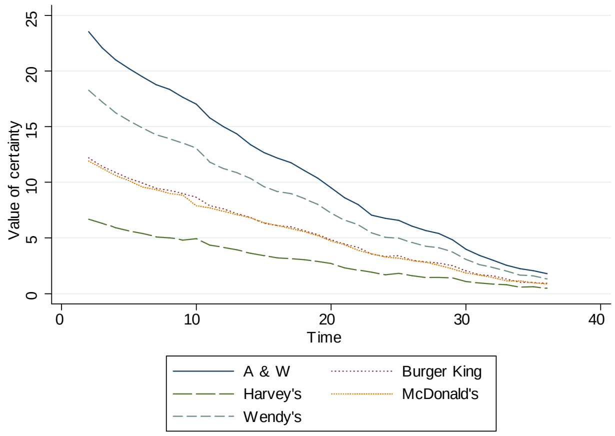


Table 19: Value of information as a proportion of monopolist's value in a local market of median size.

Chain	Proportion
A & W	4.6%
Burger King	3.3%
Harvey's	2.3%
McDonald's	2.3%
Wendy's	4.8%

Notice that the premium firms place on certainty is in general larger than their entry costs (Figure 7). Indeed, *ex ante* uncertainty can act as a significant entry barrier since firms would be better off without it. In fact, the premium they place on certainty constitutes a non-negligible portion of the net present values associated with being a monopolist (Table 18). Some noticeable patterns emerge over time. The first pattern is that the value associated with uncertainty reduction decreases for all of the chains over time. Chains in general become more knowledgeable about the markets with time; therefore, uncertainty is less of an entry barrier over time. The second pattern to note is that these values converge over time; in other words, chains have heterogeneous attitudes towards the value of uncertainty reduction in the early years. Finally, the results confirm that indeed chains value the reduction of uncertainty differently, and that Harvey's places the least value on certainty. It is well known among industry experts and academics that fast food chains invest a lot into real estate research, some more than others. The benefits associated with reducing uncertainty may indeed justify the costs associated with pre-entry real estate research.

Table 20: Relationship between the value of uncertainty reduction and market characteristics.

	(1) A & W	(2) Burger King	(3) Harvey's	(4) McDonald's	(5) Wendy's
log(Income)	0.00272***	0.00228***	0.00125***	0.000155	0.00300***
	(0.000636)	(0.000323)	(0.000211)	(0.000422)	(0.000467)
log(Population)	0.0118***	0.00589***	0.00443***	-0.00229*	0.00468**
	(0.00253)	(0.00110)	(0.000699)	(0.00102)	(0.00162)
log(Property value)	0.00488***	0.00144***	-0.000914***	0.00185***	0.00320***
	(0.000458)	(0.000246)	(0.000169)	(0.000307)	(0.000347)
log(Population density)	-0.0176***	-0.00897***	-0.00639***	0.00138	-0.00852***
	(0.00271)	(0.00118)	(0.000749)	(0.00111)	(0.00175)
Proportion work in same FSA	-0.0217	-0.0181**	-0.00941*	0.0122	-0.0264**
	(0.0119)	(0.00650)	(0.00399)	(0.00652)	(0.00940)
Own outlets in same city	-0.000101***	-0.0000582***	-0.0000216***	-0.00000980***	-0.000128***
	(0.00000164)	(0.00000104)	(0.000000542)	(0.000000275)	(0.00000207)
Constant	-0.0402***	-0.0141*	0.0115**	-0.00168	-0.0239**
	(0.0115)	(0.00605)	(0.00396)	(0.00665)	(0.00867)
Observations	16464	18341	18182	12019	18889
R^2	0.2633	0.2215	0.1338	0.1328	0.2388

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

7.2.4 Managerial implications: Geographic variation in the certainty premium

There is variation across markets in the value associated with uncertainty reduction. To what extent is this variation related to differences in characteristics across markets? Markets for which this value is high will also be markets where learning is most relevant. Therefore, linking the characteristics with these premiums will help us identify markets that are most conducive to learning. I investigate these patterns by running the following fixed effects regression:

$$Value_{imt} = \alpha_i + \boldsymbol{Z}_{mt}\boldsymbol{\beta}_i + \delta_i \cdot Existing_{imt} + \eta_m + \varepsilon_{imt}. \tag{36}$$

Here, $Value_{imt}$ is the change in variable profits associated with a counterfactual reduction of uncertainty to chain i (as a potential entrant) in market m at time t and $Existing_{imt}$ is equal to the total number of existing outlets chain i has in neighboring markets within the same city. As in the previous regressions, $\boldsymbol{Z}_{mt}$ are exogenous market characteristics.

My results in Table 19 show that for all of the chains, markets with large populations, high incomes, and valuable real estate are associated with large benefits from uncertainty reduction. In contrast, population density and the proportion of those working in the same FSA have a negative relationship with this counterfactual quantity. High density areas where workers do not commute very far to work are often associated with metropolitan cities. Fast food managers may be more well acquainted with metropolitan cities that receive more exposure in the international community; as such, their familiarity with these areas precludes the need to obtain/learn new information. Interestingly, this correlation is not pronounced for McDonald's. A natural conjecture is that as a leader in real estate research, McDonald's should not value uncertainty reduction any more or less in metropolitan areas as their research efforts overcome any informational disadvantage associated with less developed areas. Taken together, these results suggest that learning will have the least value in metropolitan areas.

There is also a geographic aspect associated with the value of uncertainty reduction, as this value decreases as the number of existing neighboring outlets within the same city increases. Although my model does not account for this feature, there could be spatial correlation in market risk. With this correlation in place, a chain may already be well-informed about a particular market if it has operated outlets in nearby markets. This finding motivates a richer model with spatially correlated learning, to be pursued in future work. Nevertheless, this spatial autocorrelation does not lead me to overstate the identified learning effects, as these effects of interest will simply be larger once I control for this correlation.

Characteristics unrelated to market risk may also be spatially correlated. FSAs in large metropolitan cities may enjoy certain benefits associated with cost and managerial efficiencies, as well as demand flows from neighboring FSAs. In other words, unobserved heterogeneity may be most prevalent in large cities, and thus lead me to overstate the benefits of reducing uncertainty in FSAs associated with those cities. Although addressing this concern directly is beyond the scope of this paper, the negative correlation between the value of uncertainty reduction and population density demonstrates that the counterfactual values associated with uncertainty reduction are not disproportionately higher for FSAs in densely populated areas.

8 Conclusion

The primary objective of my paper has been to identify the driving 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, decomposition of the spillover effect using a simple differencing technique reveals that information externalities can indeed explain the observed clustering of retail chains. Structural estimation confirms that *ex ante* uncertainty matters, and counterfactual analysis confirms that it behaves like an entry barrier, especially during the industry's early years. In particular, an industry with uncertainty is less competitive than an industry with certainty initially, but becomes just as competitive over time. I also demonstrate that an industry with uncertainty will have excessive entry/exit as the chains enter markets they would have avoided had they known the true market size. This finding suggests the importance of information - estimates of the premium these chains place on certainty indicates that the value of information constitutes a non-negligible portion of their overall values. Ultimately, my research advances our collective knowledge about the role and implications of learning in the retail industry.

In future work, I plan to explore the possibility that firms learn about profitability through their own experience in similar or neighboring markets. While the current paper models learning from others, it may be worthwhile trying to separately (or simultaneously) model intra-firm learning. Furthermore, variation in the level of experience in similar markets can introduce heterogeneity in the prior beliefs, which can ultimately be identified by data.

References

[1] Aguirregabiria, V. (2009). A Method for Implementing Counterfactual Experiments in Models with Multiple Equilibria. *Economics Letters* forthcoming.

[2] Aguirregabiria, V. (2004). Pseudo maximum likelihood estimation of structural models involving fixed-point problems. *Economics Letters* 84.

[3] Aguirregabriai, V. and Ho, C-Y. (2009). A dynamic oligopoly game of the US airline industry: Estimation and policy experiments. Working paper.

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

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

[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] Blevins, J-R. (2010). Nonparametric Identification of Dynamic Games with Discrete and Continuous Choices. Working paper.

[9] Block, J., Scribner, R., and DeSalvo, K. (2004). Fast Food, Race/Ethnicity, and Income: A Geographic Analysis. *American Journal of Preventative Medicine* 27, 212-215.

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

[11] Bronnenberg, B., Dhar, S., and Dubé, J-P. (2009). Brand History, Geography, and the Persistence of Brand Shares. *Journal of Political Economy* 117, 87-115.

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

[13] Campbell, J. and Hopenhayn, H. (2005). Market Size Matters. *Journal of Industrial Economics* 53, 1-25.

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

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

[16] Chernew, M., Gowrisankaran, G., and Scanlon, D. (2008). Learning and the value of information: Evidence from health plan report cards. *Journal of Econometrics* 144, 156-174.

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

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

[19] Ching, A., Erdem, T., and Keane, M. (2011). Learning Models: An Assessment of Progress, Changes and New Developments. Working paper.

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

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

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

[23] Drewnowski, A. and Darmon, N. (2005). Food Choices and Diet Costs: an Economic Analysis. *Journal of Nutrition* 135.

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

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

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

[27] Ellickson, P. and Misra, S. (2008). Supermarket pricing strategies. *Marketing Science* 27.

[28] Ellison, G., and Glaeser, E. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy* 105, 889-927.

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

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

[31] Fogli, A. and Veldkamp, L. (2008). Nature or Nurture? Learning and the Geography of Female Labor Force Participation. Working paper.

[32] Forbes, S. (2009). When Educators are the Learners: Social Learning and the Choice of Comprehensive School Reform Providers. Working paper.

[33] Goldfarb, A. and Xiao, M. (2011). Who thinks about the competition? Managerial ability and strategic entry in US local telephone markets. *American Economic Review* forthcoming.

[34] 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.

[35] Guiso, L. and Schivardi, F. (2007). Spillovers in Industrial Districts. *The Economic Journal* 117, 68-93.

[36] Han, L. and Hong, S-H. (2008). Testing cost inefficiency under free entry in the real estate brokerage industry. Working paper.

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

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

[39] Ivkovic, Z. and Weisbenner, S. (2007). Information Diffusion Effects in Individual Investors' Common Stock Purchases: Covet Thy Neighbors' Investment Choices. *The Review of Financial Studies* 20, 1327-1357.

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

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

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

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

[44] Kohler, H-P., Behrman, J., and Watkins, S. (2007). Social Networks and HIV/AIDS Risk Perceptions. *Demography* 44, 1-33.

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

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

[47] Lenzo, J. Competition and Complementary Activities: Lessons from Radiological Diagnosis and Treatment. (2010). Working paper.

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

[49] Magesan, A. (2010). Human Rights Treaty Ratification of Aid Receiving Countries. Working paper.

[50] Manski, C. (2004). Social Learning from Private Experiences: The Dynamics of the Selection Problem. *The Review of Economic Studies* 71, 443-459.

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

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

[53] Powell, L., Chaloupka, F., and Bao, Y. (2007). The Availability of Fast-Food and Full-Service Restaurants in the United States: Associations with Neighborhood Characteristics. *American Journal of Preventative Medicine* 33, 240-245.

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

[55] Rosenthal, S. and Strange, W. (2004). Evidence on the Nature and Sources of Agglomeration Economies. *Handbook of Regional and Urban Economics* 4. Edited by J.V. Henderson and J.F. Thisse. Elsevier.

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

[57] Schmalensee, R. (1978). Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry. *Bell Journal of Economics* 9.

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

[59] Sen, B., Shin, J., and Sudhir, K. (2011). Demand Externalities from Co-Location: Evidence from a Natural Experiment. Working paper.

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

[61] Stone, D-F. and Zafar, B. (2011). Bayesian Social Learning, Conformity or Stubbornness? Working paper.

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

[63] Sutton, J. (1991). *Sunk Costs and Market Structure*. MIT Press: Cambridge.

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

[65] Sweeting, A. and Roberts, J. Entry and Selection in Auctions. (2010). Working paper.

[66] Tauchen, G. (1986). Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions. *Economics Letters* 20, 177-81.

[67] Thomadsen, R. (2005). The Effect of Ownership Structure on Prices in Geographically Differentiated Industries. *RAND Journal of Economics* 36.

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

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

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

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

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

[73] Tucker, C., Zhang, J., and Zhu, T. (2009). Days on Market and Home Sales. NET Institute working paper.

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

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

[76] Walrath, M. Religion as an industry: estimating a strategic entry model for churches. (2008). Working paper.

[77] Ward, M. (2010). Learning to Surf: Spillover effects in the Adoption of the Internet. Working paper.

[78] Wooldridge, J. (2005). Simple solutions to the initial conditions problem for dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* 20.

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

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

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

9 Appendix: Data

9.1 List of CMA's in sample by province

All CMA's contain an urban core, which has a population of at least 100,000. Once an area has been designated as a CMA, its status will not change even if its population falls below the threshold. The set of CMA's is quite representative of Canada's urban population, as nearly all of the provinces are represented. However, because a large proportion of Canadians reside in Ontario, my sample contains a disproportionately large number of CMA's from that province.

1. Alberta: Calgary, Edmonton.
2. British Columbia: Vancouver, Victoria, Kelowna, Abbotsford.
3. Manitoba: Winnipeg.
4. New Brunswick: Moncton, Saint John.
5. Newfoundland: St. John's.
6. Nova Scotia: Halifax.
7. Ontario: Toronto, Ottawa, Hamilton, London, Windsor, Niagra Falls, Kingston, Peterborough, Guelph, Kitchener, Oshawa, Barrie, Brantford, Sudbury, Thunder Bay.

8. Quebec: Montreal, Quebec City, Sherbrooke (Missing), Saguenay (Missing), Trois-Riveres (Missing).

9. Saskatchewan: Saskatoon, Regina.

9.2 Summary of interviews with fast food executives[51]

A typical sequence of events that lead to a store opening is as follows: 1) A land developer will pitch a location to the chain; 2) a research associate assigned to the geographic region associated with the location in question will then conduct research for that market, and come up with a revenue forecast and assessment for that prospective market; and finally, 3) the head of development in that city will then give the final approval. When deciding whether to approve a location or not, they consider very carefully that location's population, number of daytime employees, and nearby shopping centres and retailers. Surprisingly, the market's average income does not seem to play a large role in McDonald's entry decisions.

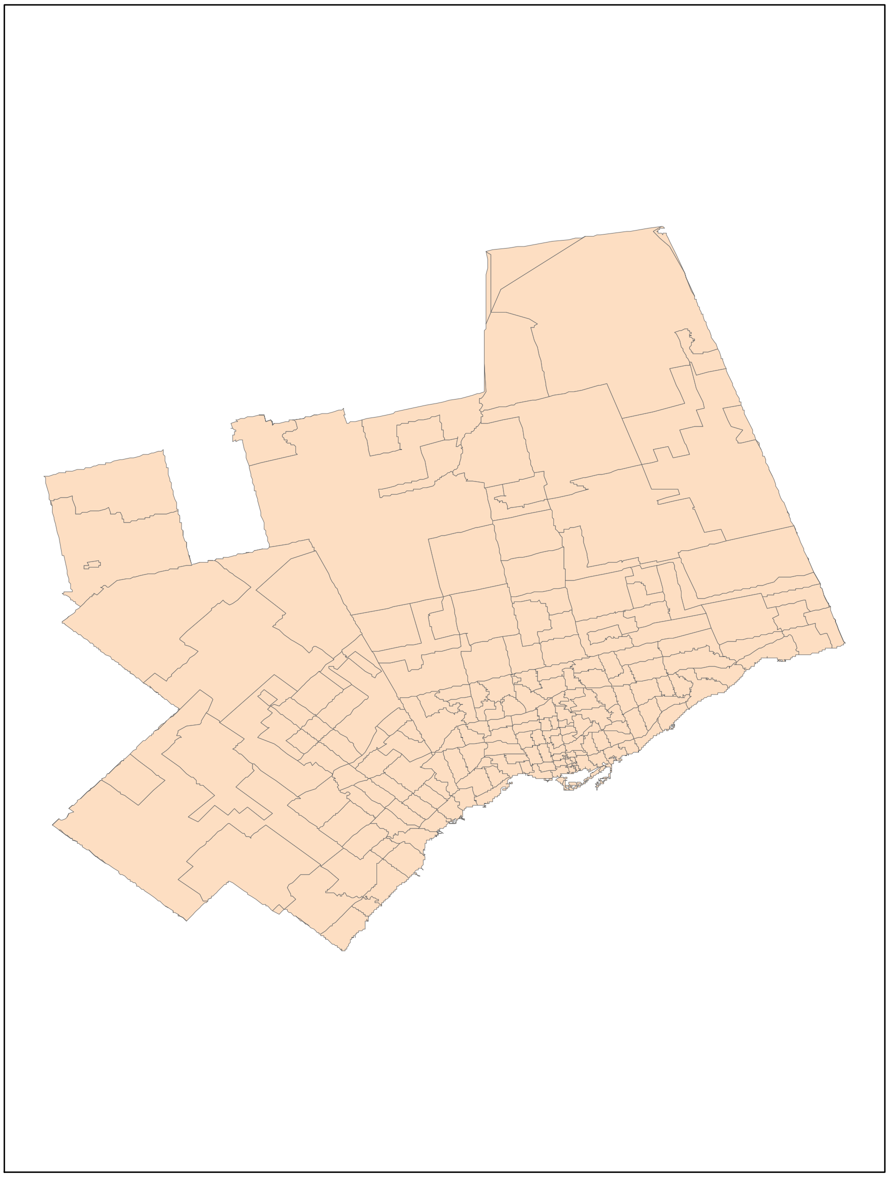
At least with McDonald's, there appears to be geographic specialization within its real estate teams. For example, one employee may be in charge of location decisions for a subregion of Toronto. For both chains, market research prior to entry is important. Researchers are well acquainted with population, income and demographic data. They also pay attention to (pedestrian) traffic flows. As Mr. White indicated, a researcher may take note of which direction pedestrians typically walk to work, and which direction pedestrians typically come back home from work. These patterns will affect their entry decisions depending on how important they view the breakfast and lunch/dinner markets. A strategy unique to Wendy's is to send researchers to an existing rival's outlet to count the number of patrons that walk into the store. This number will give them a rough idea about the demand. While each chain views one another as a competitor, they will typically enter markets that are large enough to sustain multiple chains. The chains take risk seriously. There are two main components to risk, listed in the order of their importance:

1. *Demand risk*: Associated with sales and volume. Will their franchisees be able to sell enough product to recover the costs associated with entry?

2. *Cost risk*: The physical land may not be conducive to building construction. One example would be if soil conditions are poor, or if commodity prices fluctuate a lot.

[51] The phone interview with Mark White of Wendy's was conducted on October 22, 2009 (10:00 am), and the phone interview with Patricia Simiele was conducted on November 22, 2010 (10:00 am). A face-to-face meeting with Patricia Simiele was conducted at the McDonald's Canadian headquarters (2 McDonald's Place, Toronto) on March 31, 2011 at 10:30 am. A phone interview was also conducted with Melissa Pannozzo of Harvey's on July 18, 2011.

Figure 8: Map of Forward Sortation Area (FSA) subdivisions for the city of Toronto.



10 Appendix: Model of entry/exit with learning

10.1 Numerical example

For the numerical analysis, I will be looking at a simple case with $J = 2$ chains. Entry costs are set to be the same for both of the chains, at $EC_i = EC_j = 0.1$, while their fixed costs are set to be zero ($FC_i = FC_j = 0$). The true market size is set to be $S^*_{mt} = 2$ for all m and t. I set $\theta_{1i} = \theta_{1j} = \theta_1 = 1$ and $\theta_{2i} = \theta_{2j} = \theta_2 = 0.5$. They begin with a $\lambda_0 = 0.5$ prior probability of having an incorrect assessment of market size. To solve the dynamic game under the parametrization above, I obtain the Markov Perfect Equilibrium probabilities using a policy iteration approach under the Aguirregabiria and Mira (2007) representation. As the posterior beliefs fall on the unit interval, I discretize them over a unit interval.

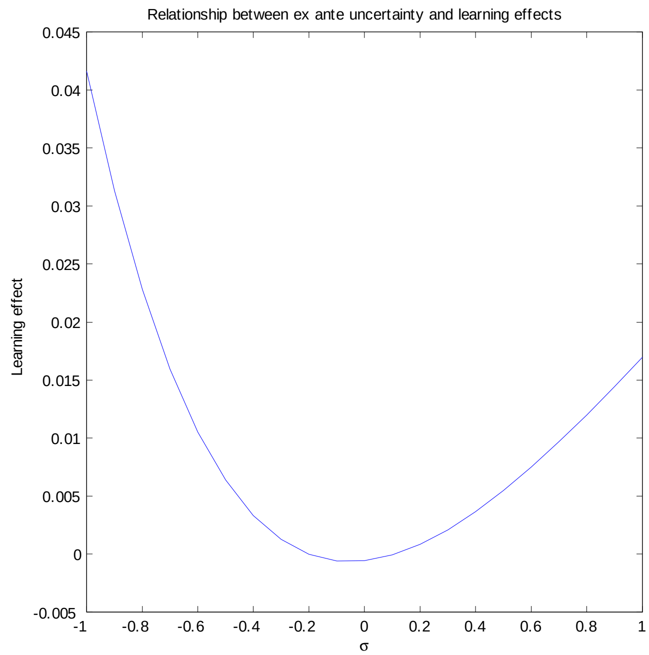
This dynamic game may have multiple Markov Perfect equilibria. In this numerical example, I have selected the (stable) equilibrium that I converge to when using policy iterations initialized with a vector of choice probabilities with all the probabilities equal to 0.5. This selected equilibrium has a non-negligible dominion of attraction in the sense that I converge to the same equilibrium when I initialize the policy iteration algorithm with vectors of choice probabilities very different to 0.5.

10.1.1 Role of uncertainty in learning

I will now explore the role of *ex ante* uncertainty in learning. Intuitively, one would expect a potential entrant facing significant uncertainty to benefit from learning. In many empirical studies on social learning, preciseness of priors are often used to justify the presence of learning. To some extent, this section validates this idea of using prior precision as a means for reduced form identification. Most importantly, this section validates a statistic that I will rely on afterwards for identification of learning. The most natural comparative static to consider would be to look at how the learning effect changes with the degree of *ex ante* uncertainty. Under the context of my theoretical model, I first define the following objects

$$\begin{aligned} P(0,1) &= \Pr(a_{it} = 1 | a_{it-1} = 0, a_{jt-1} = 1, a_{jt-2} = 1, \lambda_{t-1} \neq 0) \\ P(0,0) &= \Pr(a_{it} = 1 | a_{it-1} = 0, a_{jt-1} = 0, a_{jt-2} = 1, \lambda_{t-1} \neq 0) \\ P(1,1) &= \Pr(a_{it} = 1 | a_{it-1} = 1, a_{jt-1} = 1, a_{jt-2} = 1, \lambda_{t-1} = 0) \\ P(1,0) &= \Pr(a_{it} = 1 | a_{it-1} = 1, a_{jt-1} = 0, a_{jt-2} = 1, \lambda_{t-1} = 0). \end{aligned} \tag{37}$$

Figure 9: How does the degree of uncertainty affect the learning effect? The parameters used are $\theta_1 = 1$, $\theta_2 = 0.5$, $EC = 0.1$, $S^* = 2$ and $\lambda_0 = 0.5$.



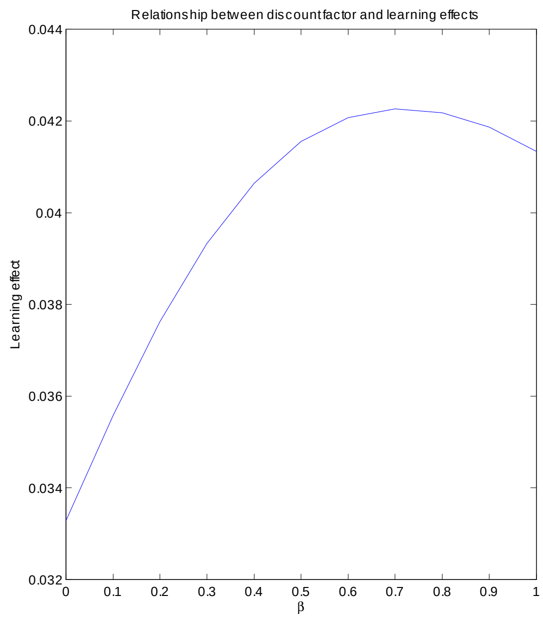
The first case describes a scenario in which i is a potential entrant that has observed j stay in the market at time $t-1$. The condition $\lambda_{t-1} \neq 0$ ensures that i did not already learn the true market size in the past. In contrast, the second case has the potential entrant observing j exiting the market. The third scenario has incumbent i observing the decision of j to stay in the market. As i is an incumbent, it has correct beliefs about the market size, as indicated by $\lambda_{t-1} = 0$. Finally, the last case has incumbent i observing j's decision to leave the market.

I now argue that the difference in conditional probabilities $\{P(0,1)-P(0,0)\}-\{P(1,1)-P(1,0)\}$ behaves like a learning effect; showing this is worthwhile as this quantity will be used in the next section to identify learning. Figure 5 illustrates the intuitive result. As $|\sigma|$ increases, the learning effect also increases. In particular, the effect is close to or slightly less than zero when $\sigma \in (-0.2, 0.2)$ and positive when $\sigma \in [-1, -0.2) \cup (0.2, 1]$. It can be as high as 0.04 when $\sigma < 0$ and 0.015 when $\sigma > 0$. These numbers seem reasonable, as my earlier estimates of learning have similar values.

10.1.2 Strategic delay

Standard models of learning in the spirit of Caplin and Leahy (1998) and Chamley and Gale (1994) demonstrate that under the presence of information externalities, agents have an incentive

Figure 10: How does the degree of forward looking affect the learning effect? The parameters used are $\theta_1 = 1$, $\theta_2 = 0.5$, $EC = 0.1$, $S^* = 2$, $\sigma = -1$ and $\lambda_0 = 0.5$.

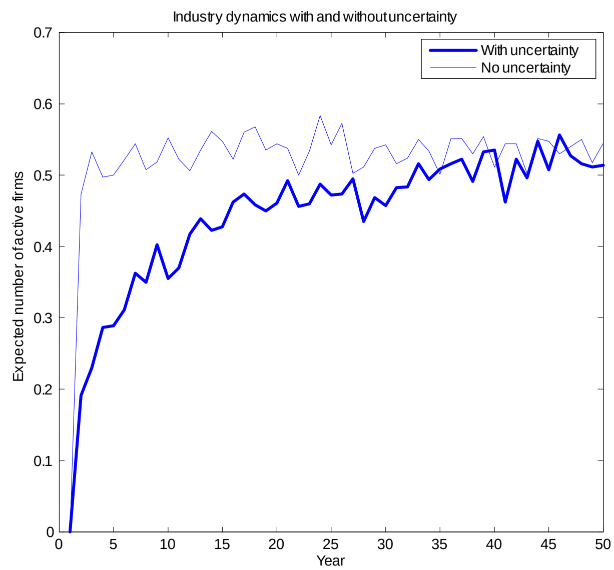


to delay their actions so as to take full advantage of these externalities. Given that my model allows for forward looking behavior, a potential entrant has an option value associated with waiting. By waiting, the potential entrant can obtain more information from the incumbent's exit/stay decisions and therefore, refine its posterior beliefs. These incentives should surface through the learning effect.

I investigate strategic delay by running comparative statics on the discount factor, β. The exercise in the previous section is repeated, except now I plot the changes in the learning effect with respect to $\beta \in (0, 1)$. Similar to the previous section, $\theta_1 = 1$, $\theta_2 = 0.5$, $EC = 0.1$, $S^* = 2$, $\sigma = -1$ and $\lambda_0 = 0.5$.

We see that for $\beta \in (0, 0.6)$, the learning effect is increasing. In the static case, firms receive no future benefit from entering markets; so continuation values may increase the probability of entry, and hence, increase the learning effect. However, for $\beta \in (0.6, 1)$, the learning effect is actually decreasing. As I make future payoffs more important, I am essentially increasing the option value of waiting at the same time. Therefore, the incentive to enter immediately following an incumbent rival's exit/stay decision is weighed against the future payoff conditional on more information. The result pertaining to the region $\beta \in (0.6, 1)$ is perhaps more applicable, as discount factors in dynamic discrete choice models are typically calibrated at 0.95 or above.

Figure 11: How does uncertainty affect industry dynamics? The parameters used are $\theta_1 = 1$, $\theta_2 = 0.5$, $EC = 0.1$, $S^* = 2$ and $\lambda_0 = 0.5$. For the case with no uncertainty, $\sigma = 0$, and the case with uncertainty has $\sigma = -1$. The expected number of active firms is defined as $\frac{1}{500}\sum_m^{500}(a_{imt} + a_{jmt})$ for a given time t.



10.1.3 Implications of learning on industry dynamics

What do industry dynamics look like with or without information externalities? In particular, can an industry's growth be accelerated by learning from competitors? I explore these questions by running a simulation using the model.

The two cases I explore are when firms face uncertainty about markets ($\sigma = -1$) or when firms face no uncertainty about markets ($\sigma = 0$). Remaining parameters are set at the same values as in the previous comparative static exercise. To perform the simulation, I randomly generate the i.i.d. extreme value shocks ε_{it} for 500 markets. Combined with the numerical solution to the dynamic game, I can generate a sequence of active/not decisions for the firms across all markets over the course of 50 years. I assume that in the first year, there are no firms active in any of the markets. These sequences are then used to tabulate the total number of markets with active firms over time.

When firms face no uncertainty about the markets, the expected number of active firms immediately goes up to about 0.5 after the first year, and fluctuates around that level over time. A very different picture emerges when uncertainty is introduced. Although this quantity converges towards 0.5 in the long run[52] under the scenario with learning, it does not happen immediately. This result

[52] I use the word long run loosely, as one may expect learning to have a level effect on the long run equilibrium. That

is similar to Rob (1991), in which his equilibrium exhibits a diffusion process.[53] There are two possible reasons for which the model can generate diffusion. The first reason is that a potential entrant uses the observed decisions of a rival incumbent to update its beliefs about the market (i.e., *time-to-learn*). It may however take time before the updated posterior reaches some level that makes the perceived net benefit of being active positive. Alternatively, a potential entrant may delay entering markets, as entering later confers greater benefit associated with the information externality from the first entrant's stay/exit decisions (i.e., *strategic delay*).

The diffusion of firms also exhibits a logarithmic pattern. Growth is fastest initially, but then slows down as the industry matures. This pattern can be reconciled with the learning story. Initially, the two firms face uncertainty in virtually all of the markets. However, this uncertainty will be resolved either through their own entry decisions, or the past exit/stay decisions of the rival. Over time, there will be fewer and fewer uncertain markets to receive favorable signals about, thus, posterior beliefs in mature markets fluctuate very little. This result is related to one of my conditions used for identification: information externalities should only be identified in situations in which past learning has not already occurred.

I now digress by going back to the Canadian fast food data to illustrate how the simulated industry dynamics establishes a "footprint" for learning. The data allows me to plot the evolution of the expected number of active firms. Unlike the model, population across markets and time is not fixed. So I group markets based on which population quartile[54] they belong to. Figure 8 illustrates that the evolution of Canada's fast food industry also exhibits a pattern of diffusion. The model's diffusion is a bit faster though when compared with some of the subplots. In the model, the expected number is between 0.4 to 0.5 after about 30 years, while the data produces a number between 0.3 and 0.4 for the first three quartiles. One explanation that is consistent with learning is that the opportunity value of delayed entry should be larger as the number of other chains increases.[55]

said, I caution the reader from extrapolating the results for too long of a time horizon. Nevertheless, the relevance of my results for the long run may not be too important, as industries in general die out eventually.

[53] He describes the process as a scenario in which entry occurs gradually, and a long run equilibrium is eventually established, but not immediately.

[54] In particular, I use the average population across time.

[55] Five chains in the actual data versus two chains in the numerical exercise.

Figure 12: A time series plot of the expected number of active firms, $\frac{1}{608}\sum_{imt} 1(a_{imt} = 1|Population_m)$.

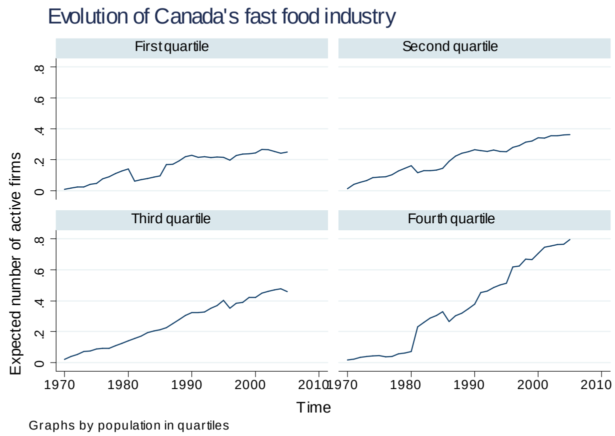


Table 21: The effect that McDonald's information externality has on the other chains.

	A & W	Burger King	Harvey's	Wendy's
McDonald's	0.02	0.01	0.00	0.00

11 Appendix: Identification of information externalities

11.1 Learning from McDonald's

My identification strategy is unable to uncover learning effects associated with past entry decisions of McDonald's. This shortcoming may be of some concern, as McDonald's is thought of as a leader in terms of real estate research and location choices. Therefore, McDonald's past entry decisions may in fact be informative. I modify the differencing estimator so as to exploit the following variation in my data: McDonald's may open more than one outlet within a particular FSA. One can then distinguish between the decision of McDonald's to add more stores, or simply stay in the market it has entered. Therefore, the potentially informative signal is from $a^{+}_{MCDmt-1} \in \{0, 1\}$, which indicates whether McDonald's opened more stores at $t-1$ after being active at time $t-2$. As before, the learning effect is identified by the same double-difference as the intuition behind the differencing remains the same, in that the first quantity captures both non-learning and learning effects, while the second quantity captures only non-learning effects.

The estimates reveal that McDonald's does exert some information externality onto others, in particular, A & W and Burger King. McDonald's information externality can increase A & W's

Table 22: The effect that McDonald's demand externalities, competition effects and unobserved heterogeneity has on the other chains.

	A & W	Burger King	Harvey's	Wendy's
McDonald's	0.01	0.00	-0.01	0.00

probability of entry by 2 percentage points, while increase Burger King's probability of entry by 1.

11.2 Calculating standard errors for double-difference decomposition

The double-difference decomposition I propose does not yield standard errors in a simple manner. To assess the statistical significance of learning, I consider a flexible random effects logit model that captures the learning effect as a parameter. To identify the learning effect, we need the following double-difference:

$$\delta = \{\tilde{V}(0,1) - \tilde{V}(0,0)\} - \{\tilde{V}(1,1) - \tilde{V}(1,0)\}. \tag{38}$$

Label $\tilde{V}(0,1) = V_i^{01}$, $\tilde{V}(0,0) = V_i^{00}$, $\tilde{V}(1,1) = V_i^{11}$, and $\tilde{V}(1,0) = V_i^{10}$. These objects can be estimated via a random effects logit estimation of the following:

$$\begin{aligned} P(a_{imt} &= 1) = \Lambda((1-a_{imt-1})(1-a_{jmt-1})V_i^{00} + (1-a_{imt-1})a_{jmt-1}V_i^{01} \\ &\quad + a_{imt-1}(1-a_{jmt-1})V_i^{10} + a_{imt-1}a_{jmt-1}V_i^{11} + \boldsymbol{Z}_{mt}\boldsymbol{\beta}_i + \eta_m) \end{aligned} \tag{39}$$

Here, η_m is unobserved market heterogeneity that I address using the 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:

$$P(a_{imt} = 1) = \Lambda(V_i^{00} + (a_{imt-1} + a_{jmt-1})(V_i^{10} - V_i^{00}) - a_{imt-1}a_{jmt-1}\delta + \boldsymbol{Z}_{mt}\boldsymbol{\beta}_i + \eta_m) \tag{40}$$

Therefore, δ can be estimated and its standard errors can be easily obtained as well. As before, we need to condition our estimation sample on unexplored markets for which exit/stay decisions of rival j are observed.

The results show that the information externalities can be precisely identified for certain chains. In particular, the estimates of δ are positive and least noisy for A & W and Burger King. In fact, the learning effects are statistically significant for A & W at the 1% level. Also note that the

Table 23: The role of information externalities using flexible logit. Here, column pertains to chain i, and row pertains to i's rival j. Note that the information externalities associated with McDonald's past stay/exit decisions is not identified as their exit is virtually negligible. Standard errors appear in brackets.

	A & W	Burger King	Harvey's	McDonald's	Wendy's
A & W	5.1 (0.53)	-0.12 (0.37)	-1.1 (0.28)	-0.22 (1.9)	-0.04 (0.42)
Burger King	4.2 (0.50)	0.30 (0.46)	-0.63 (0.34)	-0.46 (1.9)	0.38 (0.50)
Harvey's	6.9 (0.20)	0.11 (0.48)	-0.83 (0.50)	0.06 (2.0)	-0.65 (0.47)
McDonald's	N/A	N/A	N/A	N/A	N/A
Wendy's	2.9 (0.46)	0.65 (0.42)	-0.007 (0.36)	-0.06 (1.9)	-0.60 (0.60)

non-evidence for learning is precise for Harvey's, as its estimates of δ are negative and statistically significant at the 10% level. These findings are consistent with the original double-difference decomposition results, where we find no evidence of learning for Harvey's, and relatively stronger evidence of learning for A & W. Furthermore, both the signs and the statistical significance of these results are consistent with the counterfactual finding where A & W values most a reduction in uncertainty, while Harvey's values least a reduction in uncertainty.