Spatial Competition and Preemptive Entry in the Discount Retail Industry∗

Fanyin Zheng†

December 2014

Abstract

Big box retail stores have large impact on local economies and receive large subsidies from local governments. Hence it is important to understand how discount retail chains choose store locations. In this paper, I study the entry decisions of those firms, examine the role of preemptive incentives, and evaluate the impact of government subsidies on those decisions. To quantify preemptive incentives, I model firms’ entry decisions using a dynamic duopoly location game. Stores compete over the shopping-dollars of close-by consumers, making store profitability spatially interdependent. I use separability and two-stage budgeting to reduce the state space of the game and make the model tractable. Instead of adopting census geographic units, I infer market divisions from data using a clustering algorithm built on separability conditions. I introduce a ‘rolling window’ approximation to compute the value function and estimate the parameters of the game. The results suggest that preemptive incentives are important in chain stores’ location decisions and that they lead to loss of production efficiency. On average, the combined sum of current and future profits of the two firms is lowered by 1 million dollars per store. Finally, I assess the impact of government subsidies to encourage entry when one retailer exits, as happened in the recent crisis. I find that although the welfare loss such exits cause on local economies can be substantial, the average size of observed subsidies is not enough to affect firms’ entry decisions.

Keywords: Chain stores, Entry, Dynamic games, Preemption

∗I am deeply grateful to my advisors Ariel Pakes, Greg Lewis, and Guido Imbens for their constant guidance and support. I thank Jean Baccelli, Zhenyu Lai, Danial Lashkari, Robin Lee, Jing Li, Eric Maskin, Eduardo Morales, Daniel Pollmann, Marc Rysman, Mark Shepard, Che-Lin Su, Elie Tamer, Jing Xia, Lilei Xu, Ali Yurukoglu, Tom Zimmermann, and participants in the industrial organization and econometrics seminar at Harvard University for helpful comments and suggestions. Data purchased for this research is generously supported by the LEAP and Bharat Anand. For the latest version, please visit http://scholar.harvard.edu/fzheng/publications

†Department of Economics, Harvard University, 1805 Cambridge Street, Cambridge, MA 02138, USA. fzheng@fas.harvard.edu.
1 Introduction

The discount retail industry has been a fast growing sector of the U.S. economy since the 1960s. Back in 1962, only three small chains existed with less than 200 stores in total. Today, there are many more national chains with over 5000 stores in the country, generating revenue of over a hundred billion dollars per year. Such fast growth has had a large impact on local economies. On the one hand, consumers benefit from the low prices and product varieties of stores such as Walmart and Kmart. New stores of discount retailers also boost local employment. However, small businesses and other retailers suffer from the presence of discount retailers (Jia, 2008). Overall local employment rates may be lower due to entry by a discount retailer (Basker, 2007; Neumark et al., 2008). Employees criticize firms such as Walmart for driving down wages and benefits (Basker, 2007). In small towns, residents complain about dying main streets and business centers. These studies show that discount retailer’s entry has a large impact on local economies.

Multi-store retail chains are also receiving large amounts of subsidies from local governments in the form of sales tax rebate, property tax rebate, infrastructure assistance, etc. Walmart alone received over 160 million dollars in the past 15 years\(^1\). Local economies are affected as much by store closings as by store openings, as shown by Kmart closing over 1000 stores in the wake of the recent crisis\(^2\). Local governments are proposing to subsidize other retailers to replace closing stores. However, most of the abandoned retail space, such as the former Kmart stores, has remained empty for years\(^3\). Whether subsidies affect discount retailers’ entry decisions and more generally how the firms make entry decisions thus become an important question for policy makers. The first goal of this paper is to study how multi-store retail chains such as Walmart and Kmart make entry decisions.

To answer this question, I use a data set about two discount retailers. Both retailers are among the largest in the country. Since part of the data is proprietary, I am not able to reveal the identities of the firms. In what follows, I will refer to them as Blue firm and Red firm. The two firms were among the first to open discount stores in the U.S. and

\(^{1}\)Source: goodjobfirst.org.
\(^{2}\)Source: Kmart annual reports.
both experienced long periods of fast growth. Blue firm was much smaller than Red firm before the 1980s, but surpassed Red firm in the early 1990s and became one of the largest employers in the country. Blue firm succeeded in competing against Red firm because it carefully chose store locations, exploited economies of density and, most interestingly, possibly made preemptive entry moves (Bradley et al., 2002; Holmes, 2011). That is, Blue firm might have entered earlier in markets in which it feared Red firm would enter. As a consequence, the second goal of this paper is to investigate preemptive incentives and to quantify their impact on multi-store retailers’ entry decisions and on the production efficiency of opening new stores. The definition of preemptive entry I will follow hinges on how much (in equilibrium) the likelihood of one firm entering a particular location today is impacted by the likelihood of its opponent entering the same location in the future, holding its static profits constant.

The data I use consists of two major parts. The first part contains store level sales data and consumer demographic data, which will be used in the demand estimation. The second part is firms’ entry data which contains geocoded store locations and store opening dates of each Blue and Red store opened between 1985 and 2001. Blue firm’s entry data comes from Holmes (2011). I collected Red firm’s entry data using various sources including yellow page data and industry journals.

The model through which I analyze my data has two main features. First, it allows strategic interactions between firms in a dynamic duopoly framework. This is necessary because preemptive incentives cannot be studied in either a dynamic single-agent or a static game setting. Second, stores are spatially interdependent through demand and the firm-level entry decisions. This last feature fits the nature of the discount retail industry, since firms operate multiple stores that often locate close to each other. It also applies to other retail industries or markets in which firms operate in multiple interdependent locations or sectors.

As in many empirical models of dynamic games, a major obstacle to estimation is computing the value function for a large number of possible choice paths. The problem is particularly difficult to solve in the current setting given that entry decisions are made at the firm level and that decisions are not independent across markets. Therefore, I develop a series of tools to make the model tractable.

First, I apply two-stage budgeting and separability conditions to decentralize firms’
entry decisions across markets. Conditional on optimal market-level budget, if markets are separable, entry decisions become optimal within each market. This allows me to condition on the observed budget constraint of each market and solve the game for each market independently. Then, I build a clustering algorithm based on the separability conditions and apply it to partition the national market, while preserving the spatial interdependence across stores within each market. Finally, I employ a ‘rolling window’ approximation to compute value functions. That is, instead of optimizing over an infinite horizon, firms optimize over a fixed number of periods ahead and approximate the continuation value using scaled terminal values. The set of potential paths of choices each firm is optimizing over is therefore restricted, but the approximation is consistent with how managers actually make decisions. Due to the non-stationary nature of the problem at hand, the dynamic game cannot be estimated in a two-stage procedure as in (Bajari et al., 2007) (BBL) or Pakes et al. (2007) (POB). Accordingly, I solve for the nested fixed point in the estimation as in Pakes (1986) and Rust (1987). The parameter estimates are obtained by solving the game using backwards induction and maximizing the likelihood of observed location choices in each market and each period.

Using the estimated parameters, I conduct counterfactual analyses to quantify preemptive incentives and to evaluate subsidy policies. The first counterfactual analysis quantifies preemptive incentives by removing them from one firm’s optimization problem and comparing it to the original equilibrium. The challenge is that preemption is a motive instead of an action and thus it is difficult to be distinguished from other optimization motives in the entry decision. Accordingly, I use a one-period deviation approach to identify preemption. The preemptive motives of Blue firm are removed from the optimization motives of its entry decision by taking Blue firm’s observed choices out of Red firm’s choice set for one period. The reasoning is as follows: if Blue firm chose the observed locations in the current period because it feared Red firm would enter otherwise, Blue firm should delay entry at those locations now, since Red firm is not allowed to enter in the following period. Results show that preemption costs Blue firm 0.86 million dollars per store on average, which is equivalent to a small store’s one year profits. The combined current and future profits of the two firms increase by 397 million dollars when preemption is removed, which is about 1 million dollars per store. The findings thus suggest that preemptive incentives are important to multi-store retailers’ entry decisions.
and that preemptive entry can lead to substantial production efficiency loss.

In a second counterfactual analysis, I evaluate the subsidy policies proposed by local governments to encourage entry by Blue firm during a period in which Red firm exited many markets. I find that the average level of subsidies is not enough to induce entry and that preemptive incentives affect the level of subsidies Blue firm needs to enter. Finally, I compute consumer welfare loss from longer travel time to shops when a Red store closes. I find that the welfare loss can be as big as the average size of the observed subsidies Blue firm received in the past.

This paper contributes to a literature studying the discount retail industry. Holmes (2011) showed the importance of economies of scale in Walmart’s expansion, using a single-agent dynamic optimization model. Jia (2008) studied the impact of Walmart and Kmart on small business, by solving a static game between Walmart and Kmart. Ellickson et al. (2013) and Zhu and Singh (2009) also studied economies of scale and competition between big chains, in a static setting. This paper complements the literature by presenting a dynamic duopoly model to investigate the dynamic strategic interactions between firms while preserving features such as economies of scale and spatial competition that the papers mentioned above studied. In addition, the modeling and estimating methods in this paper make it possible to conduct counterfactual analyses to quantify preemptive incentives and evaluate subsidy policies.

The entry literature has been pioneered by Bresnahan and Reiss (1991) and Berry (1992). In most of the literature, for example in Mazzeo (2002) and in Seim (2006), firms make independent entry decisions in each market. In this paper by contrast, entry decisions are made at the firm level and markets are spatially interdependent. Jia (2008) allows interdependence across entry decisions, but the interdependence is assumed to be positive and linear in store density. More general forms of interdependence are allowed in this paper. They are also explicitly modeled through demand and firm level budget constraints.

This paper also contributes to a recent empirical literature on preemptive incentives. Schmidt-Dengler (2006) studied preemptive incentives in the adoption of MRI by hospitals. He identifies preemptive incentives by solving a pre-commitment game and comparing the result to the original equilibrium in which players are allowed to respond to the opponent’s action in each period. Igami and Yang (2014) examine burger chains’
preemptive entry decisions, by solving a single agent’s dynamic optimization problem and comparing the results to the dynamic duopoly equilibrium. By contrast, this paper introduces a one-period deviation method to identify preemptive incentives, while allowing static strategic interactions between firms and keeping payoffs comparable.

As for the theoretical tools used in the paper, the two-stage budgeting and separability results come from classic theorems by Gorman (1971) on consumption problems. This paper generalizes the main theorems in Gorman (1959, 1971), so that they can be applied in a dynamic game setting. The clustering algorithm developed in the paper is based on those separability results. It belongs to the class of greedy algorithms of the graph partitioning literature (Fortunato and Castellano, 2012). It is applicable to other graph partitioning or market division problems in which geographic contiguity is preserved and precision of the solution is preferred to speed.

The paper is organized as follows. Section 2 introduces the background of the industry in more detail, describes the data and provides descriptive evidence of preemptive incentives. Section 3 introduces the model, the application of two-stage budgeting and separability, and explains how markets can be defined using machine learning tools. Section 4 shows how the value functions can be approximated and presents estimation results. Counterfactuals under which preemptive motives are removed are presented in Section 5. The subsidy policy application is presented in Section 6. Section 7 concludes.

2 Industry Background and Data

2.1 Discount Retail Industry: Background

The discount retail industry in the U.S. started when Walmart and Kmart opened their first stores in 1962. It has been growing very fast in the following 40 years. The total sales of discount stores peaked at 137 billion dollars in 2001 (Census, Annual Retail Trade Survey). The discount retail industry is a very concentrated one. In 2002, the four largest firms controlled 95% of sales (Census, Economic Census).

The two firms this paper studies, Blue and Red firm, are among those four largest firms. They followed the path of growth of the industry. Blue firm has had a particularly interesting pattern of growth. It was very small at the beginning of the industry, with less than 300 stores in the 1980s when Red firm already had over 1000 stores. But it
surpassed Red firm in the 1990s and became one of the largest employers in the country. Table 1 presents the total number of stores and distribution centers of Blue and Red firm in 2001. Blue firm appears to be much bigger than Red firm in both dimensions. To explain Blue firm’s success, researchers have highlighted carefully chosen store locations, efficient distribution network, high store density, and economies of scale (Bradley et al., 2002; Holmes, 2011). Since Blue and Red firm compete in the same market, it is natural to examine whether these characteristics have also played a role in Blue firm’s surpassing Red firm. In his study of economies of scale, Holmes (2011) raises the additional question of possible preemptive entry - a topic this paper will be concerned with.

Discount retail stores are known to have a large impact on local economies. Consumers benefit from the low prices of discount stores. Ellickson and Misra (2008) find that when Walmart enters a market, its low prices extend to other local stores. Basker (2007) shows that local employment is boosted after Walmart’s entry. The impact is not always positive, however. Jia (2008) finds that half of the decline of small businesses in U.S. is caused by entry of Walmart or Kmart during the 1980s and 1990s. Basker (2007) also shows that when Walmart opens a new store, local employment shrinks in the long term due to the closings of small businesses. Because of the large and complex impact of discount retailers on the local economy, it is in the interest of policy makers to understand how decisions about where to locate stores are made - the issue this paper investigates.

Discount retailers also turn out to receive large amounts of subsidies from local governments. According to goodjobsfirst.org, Walmart alone received over 160 million dollars between 2000 and 2014. The subsidies take on various forms including sales tax rebate, property tax rebate, free land, infrastructure assistance, etc.. Since Red firm started exiting many markets in 2001, local governments have been proposing subsidies to Red firm so that it would stay or to other retailers like Blue firm so that they would enter. For example, Buffalo, NY, proposed a 400,000 dollar subsidy to Red firm for it to stays4. Lots of retail space stayed empty for years. In Rockledge, FL, for example, the ex-Red store has been empty for 11 years5. It is not clear if the proposed size of subsidies is big enough to affect retailers’ entry decisions in general - a question this paper will assess.

4Source: www.huffingtonpost.com/2012/01/26/sears-closes-cities_n_1231326.html
2.2 Data

Data limitations of the discount retail industry heavily constrains the models that can be used to analyze it. This is why I describe the data sources before presenting the model. There are four main components of the data. The first component is store and distribution center locations and time of opening between 1985 and 2003. Blue firm’s store and distribution center locations and opening dates between 1985 and 2003 come from Holmes (2011). Red firm’s locations and time of opening data come from three sources. First, addresses and time of store openings are from infoUSA in 2002. Second, I double-checked the addresses and time of opening of each store using the annual Chain Store Guide between 1984 and 2001. This step was necessary because there are 96 Red closings after 2000, and some of the stores are missing in the 2002 InfoUSA data. I do not model store closing decisions in this paper, but in the policy application in Section 6, I will discuss entry after a store closure. The time of opening and closing of these missing stores was collected by searching through local newspapers. Finally, I geocoded store addresses using the ArcGIS North America Address Locator. The distribution center addresses of Red firm have been collected from data published by the U.S. Environmental Protection Agency (EPA). The addresses have also been geocoded using ArcGIS, and opening dates have been collected from local newspapers.

Figure 1 presents the store and distribution centers of Blue firm on the map of contiguous U.S., as a snapshot by the end of 2001. The blue dots indicate Blue stores and the green diamonds indicate Blue distribution centers. Figure 2 presents the stores and distribution centers of Red firm by 2001. Each red dot is a Red store and each yellow diamond is a Red distribution center. Comparing the two maps, it appears that Blue firm has both more stores and more distribution centers, while Red stores seem to be more concentrated geographically. The figures also show that both firms are national chains and they compete in many local markets across the nation.

The sample consists of Blue and Red store openings between 1995 and 2001. 1984 and 1140 Blue and Red stores opened in this period, respectively. Store openings between

---

7 For the 12 of Red stores that I could not find information about, I assumed the time of opening to be the first quarter of the year it first appeared in Chain Store Guide, and the time of closing to be the first quarter of the year they first disappeared.
8 Distribution centers are EPA regulated facilities.
2002 and 2003 are left out of the sample because Red firm stopped opening new stores in 2002. Figures 3 and 4 display the sample store openings by year. It appears that Blue firm opened more stores than Red firm in almost every year\(^9\).

Table 2 provides the summary statistics of the characteristics of the sample by firm. The characteristics are measured for the median store in 2001. First of all, it appears that the median distance to the closest competitor’s store for Blue stores, 8.38 miles, is much bigger than for Red ones, 3.46 miles. The difference suggests that Red stores face more competition from Blue stores than Blue ones do from Red ones. Comparing this difference to the smaller difference between Blue’s median distance to the closest Blue firm, 11.90 miles, and that of Red, 10.16 miles, it appears that Blue stores are more spread out than Red stores. The number of any stores within 30 miles and population density around stores also indicate that Red stores are located in more concentrated areas, in terms of both store density and population density. Finally, Blue firm has 35 distribution centers while Red has 18. With more distribution centers, Blue stores are on average about 40 miles closer to their own distribution centers than Red stores to theirs. These differences, as will be discussed later, are important in characterizing preemptive entry behavior.

The second component of the data is store level characteristics. The store level sales estimates and square footage of selling space of Blue and Red firm in 2007 come from the Nielsen TDLinx data. The sales are estimated using multiple sources including self-reported retailer input, store visits, questionnaires to store managers, etc.. They are regarded as the best available store level sales data of the discount retail industry and as a consequence they have been used by other researchers (Ellickson et al., 2013, Holmes, 2011). For stores that sell both general merchandise and grocery, only sales of general merchandise are included. Square footage of selling space is derived from actual property plans. Because of the proprietary nature of this data set, I cannot present summary statistics of store characteristics.

The third component of the data consists of demographic information, wage, rent, and other information about the two firms. I use block group level demographic data in the 1980, 1990, and 2000 decennial censuses. A block group is a geographic unit that has a

\(^9\)The peak for Red firm in 1992 corresponds to the acquisition of a small chain. The stores belonging to the small chain are not counted as entry but kept in the sample as “Red stores” after the acquisition.
population between 600 to 3000 people. The demographic information of each block group contains total population, per capita income, share of African-American population, share of elderly population (65 years old and above), and share of young population (21 and below). Table 3 presents the summary statistics of the block group level demographic information. Wage data is constructed using average retail wage by county in the County Business Patterns between 1985 and 2003. Rent data is created using the residential property value information in the 1980, 1990, and 2000 decennial censuses. I adopt the same method as in Holmes (2011) to construct an index of property values. (See Appendix A of Holmes (2011) for details.) Firms’ annual reports and interviews I conducted with managers and consultants also provide supporting information. Goodjobfirst.org is a website that collects government subsidy data published from various sources. The list of subsidies are incomplete, but it gives an idea about the scale of the subsidies. It is the best data source of its kind. Shoag and Veuger (2014) use this data in their study. I also interviewed a manager of Blue firm and a former manager of Red firm. I use the information collected from those interviews to choose between different modeling options, so that the model mimics how managers make decisions in reality.

2.3 Descriptive Evidence of Preemptive Entry

In this section, I provide suggestive evidence of preemptive entry using reduced form regressions. Preemption, in this context, refers to the entry by one firm in order to deter entry by its opponent. More specifically, I define it as how much, in equilibrium, the likelihood of one firm entering a particular location today is impacted by the likelihood of its opponent entering the same location in the future, holding static profits constant. (A formal definition will be given in Section 5.) Using descriptive data, three kinds of behavior could be called preemption. First, a firm can open more stores than otherwise optimal, so that the opponent cannot enter the market. Second, a firm can cluster its stores to deter entry of the competitor. Finally, a firm can open a store earlier than otherwise optimal, so that the competitor cannot enter. The first two types are hard to find evidence for using reduced form regressions, since it is hard to separate store quantity and store density from unobserved market profitability. Therefore, I focus on the timing of store opening. More precisely, I choose to study Blue firm’s store opening
time instead of Red firm’s, for two reasons. First, Blue firm’s fast growth and high
store density suggests it is more likely that it engaged in preemption, as described in
the previous section. Second, during the observed period of time, Blue firm has more
observations of new stores openings than Red firm.

Next, I describe how preemptive incentives can be identified. The goal is to find
a location characteristic that 1) affects Red firm’s payoff of entering the location and
therefore Blue firm’s dynamic payoff if Blue firm does not enter the location in the
current period, 2) does not affect Blue store’s static profits of entry in the current period,
i.e. is not correlated with unobserved market profitability. In other words, the impact of
this location characteristic on Blue store’s opening time should indicate how much the
likelihood of Red store’s entry affects Blue firm’s entry decision.

One variable satisfying these conditions is the distance between a Blue firm’s store
and the closest Red firm’s distribution center. The distance to the Red distribution
center affects Red firm’s payoff of locating a store, thus the likelihood of Red firm’s
entry. On the other hand, this distance does not directly impact Blue firm’s static
profit of entry. The challenge is that distribution centers are likely to be located close
to potential stores, so that locations of Red distribution centers can be correlated with
unobserved market profitability in the area. I include in the regression a control variable
that approximates the unobserved market profitability around each Red distribution
center. The profitability is measured by the total number of stores around the Red
distribution center, including both Blue and Red stores, by the end of the observed
period\(^\text{10}\).

The Cox hazard model is applied to examine the impact of the distance to Red
distribution center on Blue store’s opening time. The dependent variable is duration
before store opening for each Blue store \(l\), measured in quarter. The observed time period
is between 1985 and 2001. The independent variable of interest is the distance between
\(l\) and the closest Red distribution center. Since Red firm was expanding its distribution
center network during the period of observation, the distance to Red distribution center
is time dependent. Thus each observation is a location \(l\) observed in period \(t\). Let \(h_{lt}\) be

\(^{10}\text{The underlying assumption in the analysis is that the unobserved market profitability that is corre-
related with the locations of distribution centers does not fluctuate very much over time. This is likely
to be true since all distribution centers are located in very rural and remote areas where demographics
did not change very much over the sample period.}
the store opening hazard rate of location \( l \) in period \( t \).

\[
\ln(h_{lt}) = \ln(h_{0t}) + \beta_1 d_l(t) + \beta_2 x'_l(t),
\]

where \( h_{0t}(t) \) is the baseline hazard rate at time \( t \), \( d_l(t) \) is the distance between location \( l \) and its closest Red distribution center, and \( x'_l(t) \) is a set of control variables. The control variables include Blue firm’s store and distribution center network characteristics, Red firm’s store characteristics and other location characteristics such as wage, rent, and demographics. (See table 4 for detailed descriptions.) Since the stores in the sample are those that did eventually get opened, the regression captures the incentives for firms to manipulate the order of actions for strategic reasons.

Column 1 of Table 5 presents the results of the Cox hazard regression. Standard errors are clustered at the location level. The estimate indicates that a 100 mile increase in the distance between a Blue store and the closest Red distribution center reduces the hazard rate of Blue firm’s store opening by 1.6%. Table 5 column 2 reports the same regression using an OLS framework. In this case, each observation is a store location. The estimated coefficient on distance to Red distribution shows that when the distance between a Blue store and its closest Red distribution center decreases by 100 miles, the opening time of the store becomes 1.2 quarters earlier on average. These results suggest that, if Red firm is also more likely to enter the same location, Blue firm is more likely to enter the location earlier than otherwise. This is suggestive evidence of preemption.

3 Model

3.1 Overview

The model consists of two parts, a demand model and an entry model. The demand model is needed for computing sales of each existing and potential store. It includes detailed geographic information of consumer and store locations which allows store sales to be spatially interdependent. Another source of spatial interdependence across stores comes from the entry model. Entry decisions are modeled at the firm level subject to a budget constraint in a dynamic discrete-choice game framework. Then two-stage

\[11\] Since Blue firm was also expanding its distribution center network, distance to Blue’s distribution centers is also time dependent.
budgeting and separability are applied to make the model tractable. Finally, a clustering algorithm based on separability conditions is employed to define markets. Section 3.2 describes the demand model, and Section 3.3 explains the entry model.

3.2 Demand

A demand model is needed in order to compute sales for each store location given consumer demographic information and location characteristics. There are two main ways to model demand in the literature. First, one can follow a Berry et al. (1995) type of model in which consumers in each market choose from the same set of products. Markets are independent and heterogeneity in consumer characteristics translates to different market shares of the same product in each market. This model allows for unobserved preference heterogeneity via random coefficients. Alternatively, one can adopt the demand model as in Holmes (2011). In this model, there is no market division and each consumer has its own choice set. This model has detailed geographic information about consumers and stores, and generates spatial interdependence across store locations. The drawback of this model is that it does not allow for unobserved preference heterogeneity due to the burden of computing a different set of choice probabilities for each consumer.

I choose the latter approach because spatial interdependence is important for modeling chain store’s entry decision. Since payoffs are maximized at the firm level, when evaluating the payoff of a new store, firms need to take into account the impact of existing stores, including each firm’s own stores and its competitor’s stores. For example, if Blue firm is considering opening a new store in Boston, MA, it needs to evaluate the profitability of this location bearing in mind the existing stores in the neighboring town of Cambridge, since people living in both Boston and Cambridge can easily shop from all the stores located in either of the two cities.

The drawback of this approach is that it does not allow for unobserved preference heterogeneity. For example, the model is not able to capture the fact that different consumers dislike distance between home and stores with different intensities. In theory, random coefficients can be added to the model to account for unobserved heterogeneity. In practice, however, it is computationally infeasible. The model is already difficult to estimate given that each unit of consumers (i.e. a block group) has a different choice set
and a different set of choice probabilities and that there are over 200,000 block groups in the continental U.S. However, interaction terms in the regression and the definition of choice set for each block group can be used to mediate the problem. A detailed explanation is provided later in this section.

Each consumer $i$ is a block group. Let $u_{ijl}$ be the utility of consumer $i$ shopping at firm $j$’s store $l$.

$$u_{ijl} = \beta x_{jl} + \gamma_1 d_{il} + \gamma_2 d_{il} \times popden_i + \varepsilon_{ijl},$$

where $x_{jl}$ is a vector of store characteristics including a brand dummy indicating if the store belongs to Blue or Red firm, the size of the store and if the store is newly opened in the current period. $d_{il}$ is the distance between consumer $i$ and store $l$. $popden_i$ is population density at block group $i$. Population density is measured by log of thousand people\textsuperscript{12} within 5 miles of block group $i$. When population density varies, the interaction of distance and population density captures the heterogeneity in consumers’ preferences with respect to distance to shops. $\varepsilon_{ijl}$’s are independent identically distributed and follow a type I extreme value distribution. Let $u_{i0}$ be consumer $i$’s utility of shopping from an outside option, i.e. a store that does not belong to either Blue or Red firm:

$$u_{i0} = \alpha w_i + \varepsilon_{i0},$$

where $w_i$ is a vector of covariates that include a constant, population density, population density squared, per capita income, and share of african-american, elderly, and young in the population. $u_{i0}$ allows the utility of shopping from the outside option to depend on location characteristics. For example, more populated areas have more outside options and thus higher utility of not shopping from any Blue or Red stores in the choice set. This attempts to control for other competitors Blue and Red firms face in the market.

Block group $i$’s choice set is defined as Blue and Red stores within $r_i$ miles of the block group. $r_i$ is a function of population density, $25 \times (1 + (\text{median}(popden) - popden_i)/\text{median}(popden))$. $r_i$ is in the interval between 17 and 35 miles, and equals 25 miles for the median block group with respect to population density. Letting $r_i$ depend on population density captures the heterogeneity of consumer preferences towards distance between home and shop across areas with different population density, in terms of

\textsuperscript{12}Block groups with less than 1000 people are grouped together.
of the furthest store they are willing to travel to. $r_i$ increases as population density decreases. In other words, people living in rural areas might be willing to travel further to a shop than those living in urban areas.

Let $p_{ijl}$ be the probability of consumer $i$ shopping at store $l$. Then store $l$'s revenue is

$$R_{jl} = \sum_{i: d_{il} \leq r_i} \lambda \cdot p_{ijl} \cdot n_i,$$

where $\lambda$ is average spending per consumer and $n_i$ is the total population in block group $i$. Ideally $\lambda$ might depend on consumer characteristics $w_i$. But the data is not detailed enough to identify $\lambda(w_i)$. This is because sales are only observed at the store level and that each store has a different set of consumers $i$ patronizing it. One would need individual consumer level spending data to identify $\lambda(w_i)$. The constant $\lambda$ is the average spending per consumer across the nation. In one of the empirical specifications, $\lambda$ is allowed to depend on if store $j$ sells general merchandize only or both general merchandize and grocery\(^{13}\). Results do not change very much. (See Section 4.1 for details.)

### 3.3 Firm’s entry decision

In this section, I describe the firm entry model and show how it becomes tractable by applying two-stage budgeting and a clustering algorithm. I give an overview of the model in 3.3.1, present the details of the model in 3.3.2, discuss two-stage budgeting in 3.3.3, derive the separability conditions in 3.3.4, and explain the clustering algorithm in 3.3.5.

#### 3.3.1 Overview of multi-store chain’s entry model

There are three features of this multi-store chain’s entry model. The first feature is that firms maximize payoffs over all stores instead of at each store independently. This is important because of the nature of multi-store retail chains, as well as of the spatial interdependence between stores as illustrated by the demand estimation in 4.1. The second feature is that firms are forward-looking. This is necessary when examining preemptive incentives. Given that demographics and distribution networks are changing over time, it is reasonable to assume that firms maximize the sum of expected current

---

\(^{13}\)As described in Section 2.2, the sales data of stores that sell both general merchandize and grocery only includes sales of general merchandize.
and future payoffs. Holmes (2011) also showed that dynamic consideration is important for discount retail chain’s entry decisions. The third feature is that there are strategic interactions between firms. This is also necessary for studying preemptive entry. It is supported by the fact that Blue and Red firm compete in many markets as shown in Figure 1 and Figure 2. Findings in Jia (2008) provide more evidence of strategic interactions. Therefore, I model firms’ entry decisions using a discrete choice game framework in which decisions are made at the firm level.

In each period, firms choose the locations of a set of new stores to maximize the current profits and the sum of discounted future values. When making the decision, each firm takes into account the current and future demographics, distribution networks, local wage and rent, its own store openings in the future, and its opponent’s store openings in current and future periods. The decision is made at the firm level instead of individual store level. The number of new stores to be opened is determined by a budget constraint. Since both firms were expanding in the sample period, they face financial constraints. Moreover, the constraint is necessary for studying preemptive incentives. Firms move sequentially each period. Blue firm moves first.

The large number of possible locations and store openings of both firms in each period leads to the very large state space in the game. As a result, the firm optimization problem has high computational complexity for the firms as well as for the econometrician. Tools that make the model tractable mimicking the way firms solve the problem in reality are therefore desirable.

### 3.3.2 Two-player discrete choice game

Let $\pi^l_{jt}$ be the static profit of firm $j$’s store $l$ in period $t$.

$$
\pi^l_{jt} = \mu_j R^l_{jt}(s_t) - w^l_t E(R^l_{jt}(s_t)) - r^l_t L(R^l_{jt}(s_t)) - \psi_j D^l_{jt} - \alpha_j x^l_{jt},
$$

(3.2)

where $\mu_j$ is the gross margin of firm $j$. $s_t = (s_{jt}, s_{-jt})$, and $s_{jt} = \{0, 1\}^L$ indicating if $j$ has a store at each location of all possible locations $\{1, ..., L\}$. Each location $l$ contains up to one store. Denote $j$’s opponent by $-j$. $R^l_{jt}(s_t)$ is the revenue of store $l$ in period $t$ which depends on $s_t$, the locations of both $j$’s stores and $j$’s opponent’s stores. I follow Holmes (2011) in modeling labor cost and land cost as variable cost. $w^l_t$ and $r^l_t$ are local wage and rent. $E(R^l_{jt}(s_t))$ is the number of employees and $L(R^l_{jt}(s_t))$ is the size of land.
\( \psi_j D_{jt} \) is the distribution cost where \( \psi_j \) is per unit distribution cost and \( D_{jt} \) is distance to distribution center. \( \alpha_j x_{jt} \) is fixed cost which depends on population density around store \( l, x_{jt} \). Each period \( t \) is a quarter. The firm level static profit is

\[
\pi_{jt} = \sum_{l=1}^{L} s_{jt} \left\{ \mu_j R_{jt}(s_l) - w_j^l E(R_{jt}(s_l)) - r_j^l L(R_{jt}(s_l)) - \psi_j D_{jt} - \alpha_j x_{jt} \right\}, \tag{3.3}
\]

where the sum is over all the locations of firm \( j \) stores.

Next I introduce the value function of the firm. For simplicity, I assume sequential move and that Blue firm moves first\(^{14} \). \( a_{jt}^l \) denotes firm \( j \)'s action at location \( l \) in period \( t \), where \( l \in \{1 \cdots L_t\} \). \( a_{jt}^l = 1 \) if \( j \) opens a new store at \( l \) in period \( t \), and \( a_{jt}^l = 0 \) otherwise. \( L_t \) is the set of all possible locations minus those taken by the two firms before period \( t \), i.e. \( L_t = L / \{ s_{jt}, s_{-jt} \} \). \( s_{jt+1} = s_{jt} + a_{jt} \), where \( s_{jt} \in \{0, 1\}^L \). Let \( z_{jt} \) be the location of \( j \)'s distribution centers in period \( t \) and \( B_{jt} \) be the budget constraint of firm \( j \) in period \( t \). For notational simplicity, let \( s_{jt} = (s_{jt}, z_{jt}, B_{jt}) \), and \( s_t = (s_{jt}, s_{-jt}) \) be the state variable. Firm \( j \)'s value function in period \( t \) is

\[
V(s_{jt}, s_{-jt}) = \max_{a_{jt} \in A_t} \left\{ \mathbb{E} \pi(s_{jt} + a_{jt}, s_{-jt}) + \beta \sum_{s_{jt+1}} \mathbb{E} V(s_{jt} + a_{jt}, s_{jt+1}) P(s_{jt+1}|s_{jt}, s_{-jt}) \right\} \tag{3.4}
\]

s.t.

\[
\sum_{l=1}^{L_t} f(a_{jt}^l) \leq B_{jt}, \tag{3.5}
\]

where \( A_t = \{0, 1\}^{L_t} \) is the choice set in period \( t \) subject to the budget constraint \( B_{jt} \). The expectation is over a cost shock \( \eta_{jt} \) of entering at location \( l \). \( \eta_{jt} \) are i.i.d. across locations and time periods. \( P(s_{jt+1}|s_{jt}, s_{-jt}) \) is the transition probability of \( j \)'s opponent in period \( t \). \( f(a_{jt}^l) \) is the budget function which I will discuss in more detail. \( \beta \) is the discount factor.

Each period, firm \( j \) chooses the optimal entry decision \( a_{jt} \in A_t \) to maximize the sum of expected profits \( \mathbb{E} \pi(s_{jt} + a_{jt}, s_{-jt}) \) and the continuation value \( \beta \sum_{s_{jt+1}} \mathbb{E} V(s_{jt} + a_{jt}, s_{jt+1}) P(s_{jt+1}|s_{jt}, s_{-jt}) \). The distribution of \( \eta_{jt} \) is common knowledge but its

\(^{14} \)This is a strong assumption. In the future, I would like to flip the order and solve the game for when Red firm moves first.
realization is private information. Firm $j$’s strategy $\sigma_j$ is a function from the state variable $s_t$ to a set of choice probabilities $Pr(a_{jt}|s_t)$. Perception of future states $P(s_{t+1}|s_t)$ is consistent with equilibrium play. The solution is a Bayesian Markov Perfect Equilibrium.

The difference between this model and the commonly used incomplete information dynamic game framework (Ryan, 2012) is that it has a budget constraint in Equation (3.5). Think of the budget function $f$ as a cost of opening $\sum_t a'_{jt}$ stores in period $t$. It represents the actual costs of acquiring land or building infrastructure, as well as the management costs of hiring workers or submitting paperwork to the local government. There are three reasons to include this condition. First, it mimics the way firms behave. According to the managers I interviewed, each period, firms designate a certain amount of funds for the opening of new stores, which is equivalent to a budget constraint. Second, both firms are expanding in the sample period 1985-2001. Financial constraints can often be a serious consideration when firms are expanding. Figure 5 plots the book value of total assets of Blue and Red firm in the sample period, in respective colors. The figure shows that Blue firm’s book value of total assets grew fast in this period. Thus it is likely that the financial constraints Blue firm is facing are substantial at the beginning of this period and are reduced by the end of this period. Third, the budget constraints are necessary for studying preemptive incentives. If firms were not liquidity constrained, in theory, they could enter all markets to deter entry by the competitor in period 0. This is unrealistic both in terms of financial costs and management costs. For simplicity, I assume hereafter that $\sum_{i=1}^{L_t} f(a^t_{jt}) = \sum_{i=1}^{L_t} a^t_{jt} \leq B_{jt}$, i.e. the total number new stores each firm can open in each period is held fixed at the observed level. Note that although $B_{jt}$ is a choice made by the firm, the assumption does not cause selection issues since the choice probabilities become conditional probabilities given the optimal budget constraint $B_{jt}$.

Next I explain how the set of potential locations $L$ is defined in the game. I restrict $L$ to be all the locations Blue and Red firm eventually entered by the end of the sample period. The alternative would be to include all possible locations regardless of the existence of a store at any point in time. There are two reasons for choosing the former approach over the alternative one. First, for the purpose of studying preemptive incentives, it is reasonable to focus on locations firms are potentially interested in entering. If a location is very far from being profitable enough for either firm to ever enter, it does not provide
information for identifying preemptive incentives. Second, including locations where no entry is ever observed implies dividing the U.S. national market into many smaller markets. The division usually involves using census geographic units as markets (Jia, 2008, Zhu and Singh, 2009, Ellickson et al., 2013). This allows little spatial competition and, as shown below, may lead to biased results. The drawback of defining observed stores by the end of the sample period as the set of potential locations is that firm’s decisions are very much affected by the limited choice set towards the end of the sample period. To avoid the problem, I leave out the last two years of data and only include observations between 1985 and 1999 as the sample of study. The choice of leaving out two years, specifically, will be explained in 4.3.

Modeling entry decisions at firm level and allowing for spatial interdependence of store locations captures the nature of spatial competition between multi-store chains, but it also makes the problem intractable. Each period, firms are choosing from the set of potential locations that have not been occupied. Since both firms are expanding very fast in the sample period, the choice set is very large. Take \( t = 36 \), the fourth quarter of 1993, as an example, Blue firm and Red firm opened 27 and 24 stores respectively. The total number of potential locations is 1262. With a total of 1262 locations, the size of the state space is \( \binom{1262}{27} \approx 10^{35} \).

### 3.3.3 Two-stage budgeting

In this section, I describe how two-stage budgeting and separability can be applied to make the model tractable, while retaining the features described above. Two-stage budgeting refers to the fact that consumers first allocate a given amount of total expenditure to categories of goods and then optimize consumption within each category, conditional on the amount of expenditure designated to this category of goods (Gorman, 1971)\(^{15}\). I apply this idea to chain stores’ entry problem. Store locations are similar to goods in the consumption problem. Let \( \{1, ..., P_{Mt}\} \) be a partition of potential locations \( \{1, ..., L_t\} \) in period \( t \). Partitions mimic the categories of goods in the consumption problem. Two-stage budgeting implies that firms solve the following problem. For each \( P_m \in \{1, ..., P_{Mt}\}, \)

\(^{15}\)Note only the separability conditions are needed here, the conditions for constructing price index are not.
\[ V(s_{jmt}, s_{-jmt}|B_{mt}) = \max_{a_{jmt} \in \mathcal{A}} \left\{ \mathbb{E} \pi(s_{jmt} + a_{jmt}, s_{-jmt}) + \beta \sum_{s_{-jmt+1}} \mathbb{E} V(s_{jmt} + a_{jmt}, s_{-jmt+1}|B_{mt}) \cdot P(s_{-jmt+1}|s_{jmt+1}, s_{-jmt}) \right\} \quad (3.6) \]

s.t.
\[ \sum_{l \in m} a_{jt}^l \leq B_{jmt}, \]
where \( s_{jmt} = \{0,1\}^m \) and \( B_{jmt} \) is the budget constraint of element \( P_m \) of the partition \( \{1,...,P_M\} \). 3.6 corresponds to the consumption problem in stage 2. Then, firm \( j \) solves for the optimal budget \( \{B_{j1t},...,B_{jMt}\} \) for each element \( P_m \) of the partition:
\[ \sum_{m=1}^{M_t} \mathbb{E} V(s_{jmt}, s_{-jmt}|B_{jmt}) \quad (3.7) \]

s.t.
\[ \sum_{m=1}^{M_t} B_{jmt} \leq B_{jt}, \]
which corresponds to the stage 1 of the consumption problem.

With two-stage budgeting, firms solve two smaller optimization problems, (3.6) and (3.7) instead of the problem in (3.4). This decentralization greatly reduces the size of the state space in estimation. When estimating a model like (3.6), one can condition on the the optimal budget of each element of the partition \( \{B_{jmt}\}_{m=1}^{M_t} \), and only solve the stage 2 problem (3.6). The state space is then reduced to \( \sum_{m=1}^{M_t} \left( \frac{|P_m|}{\sum_{l \in m} a_{jt}^l} \right) \).

Moreover, two-stage budgeting is a good approximation to how firms actually behave. Blue firm, for example, divides the U.S. national market into regions. According to the manager I interviewed, regional managers choose a set of potential new store locations each period, and submit the locations to the headquarter. Managers in the headquarter then rank the potential locations from all regions and decide which stores will be opened subject to a budget constraint.

However, it is not clear whether solving (3.6) and (3.7) is equivalent to solving (3.4). In the next two sections, I show that under a set of conditions, namely separability,
solving the two-stage budgeting problem in (3.6) and (3.7) is equivalent to solving the overall optimization problem (3.4). Then I derive sufficient conditions on the primitives of the model such that the separability conditions are satisfied. I define a market to be an element of the partition \( P_m \). I show that the solution to the two-stage budgeting problem is optimal if separability across markets holds.

3.3.4 Separability conditions in a two-player Markov game

Separability is defined following the work of Gorman (1959, 1971) and is further generalized to be applicable to a two-player dynamic game setting. First, to build intuition, I define separability in the static game context. For simplicity, the subscript \( t \) is suppressed. Let \( \sigma_j \) be firm \( j \)'s strategy. \( s = (s_j, s_{-j}) \) is the state variable. Firms solve

\[
\max_{\sigma_j(s)} \pi(s_j + a_j, s_{-j} + a_{-j}) \text{ s.t. } \sum_{l=1}^L a_j^l \leq B_j.
\]  

(3.8)

Let \( \{P_1, \cdots, P_M\} \) be a partition of the potential store locations \( \{1, \cdots, L\} \). Denote \( \pi(s_j^l = 1, s_{-j}^l, s_{-j}) \) the profit of \( j \) when \( j \) has a store at location \( l \). Define

\[
\Delta \mathbb{E}(\pi(s_j, s_{-j}, l) = \mathbb{E}[\pi(s_j^l = 1, s_{-j}^l, s_{-j}) - \pi(s_j^l = 0, s_{-j}^l, s_{-j})],
\]

to be the expected marginal profit of \( j \) entering \( l \) when the state variable \( s \) equals \( (s_j, s_{-j}) \), where the expectation is taken over the cost shock \( \eta_j^l \).

**Definition 1** Locations \( \{1, \cdots, L\} \) are separable in the partition \( \{P_1, \cdots, P_M\} \) if

\[
\frac{\Delta \mathbb{E}(\pi(s_j, s_{-j}, l))}{\Delta \mathbb{E}(\pi(s_j, s_{-j}, h))} \perp (s_j^k, s_{-j}^k), \forall l, h \in P_m, \forall k \notin P_m, \forall l, h \in P_m, \forall k \notin P_m.
\]

where \( l, h \in P_{m_l} \), and \( k \in P_{m_k} \).

In other words, if the ratio of the expected marginal profits of opening any two locations in a market does not depend on the state variables in another market, locations are separable with respect to markets. This too is analogous to the consumption problem, in which separability holds when the rates of substitution of any two goods are independent across categories of goods (Gorman, 1959). Next, I define separability in strategy \( \sigma_j \). Note \( \sigma_j(s) \) can be written as a vector \( (\sigma_j^1(s), \cdots, \sigma_j^M(s)) \) for any partition \( \{P_1, \cdots, P_M\} \). Similarly, any state variable \( s_j \) can be written as a vector \( (s_{j1}, \cdots, s_{jM}) \). Let \( \sigma_j^* \) be the best response of \( j \) given opponent's strategy \( \sigma_{-j} \), and \( a_j^* \) be the corresponding optimal action at state \( (s_j, s_{-j}) \).
**Definition 2** Firm \( j \)'s strategy \( \sigma_j^* \) is separable in the partition \( \{P_1, \cdots, P_M\} \) if for given \( \sigma_{-j}, \exists \sigma_{j1}^*, \sigma_{j2}^*, \cdots \sigma_{jM}^* \) s.t.

\[
\sigma_{jm}^*(s_{jm}, s_{-jm}, B_m) = \sigma_{jm}^*(s_j, s_{-j}, B),
\]

where \( \sigma_j^* = (\sigma_{j1}^*, \cdots, \sigma_{jM}^*) \), \( B_m = (B_{jm}^*, B_{-jm}) \), \( B_{jm}^* = \sum_{l \in P_m} a_{jl}^l \), \( B_{-jm} = \sum_{l \in P_m} a_{jl-j}^l \), \( \forall m = 1, \cdots, M \), and \( \sum_{m=1}^{M} B_m = B \).

In other words, \( \sigma_j^* \) is separable if each of its component \( \sigma_{jm}^* \) can be written as a function \( \sigma_{jm}^* \) which only depends on the state variable in the partition \( m \), \( (s_{jm}, s_{-jm}) \), and on the budget constraint of the partition, \( B_m \). This implies that conditional on the optimal budget of the partition \( B_{jm}^* \), \( j \) is able to compute the best response in partition \( j \) with information within the partition \( m \) only, regardless of the values of state variables or budget levels in other components of the partition.

**Theorem 1** If locations \( \{1, \cdots, L\} \) are separable in partition \( \{P_1, \cdots, P_M\} \), and the opponent's strategy \( \sigma_{-j}^* \) is separable, then \( j \)'s optimal strategy \( \sigma_j^* \) is separable.

See Appendix I for the details of the proof. Theorem 1 states that if locations \( \{1, \cdots, L\} \) are separable and that one firm is playing a separable strategy, then it must be optimal for the other firm to play a separable strategy as well. In other words, both firm's strategies are separable in equilibrium. Define such an equilibrium as **separable equilibrium**. Separable equilibrium is a refinement of Nash equilibrium.

Next, I derive sufficient conditions on the primitives such that separability of locations holds. There are four parts of the profit function (3.3) that need to be examined for separability. The first three terms in the profit function all depend on revenue \( R_{jl}(s_l) \), thus \( R_{jl}(s) \) needs to satisfy the separability condition. The other three terms are the distribution cost, the fixed cost, and the cost shock when opening a new store \( \eta_j^l \).

**Theorem 2** The location \( \{1, \cdots, L\} \) is separable in partition \( \{P_1, \cdots, P_M\} \) if the profit function \( \pi(\cdot) \) satisfies the following conditions,

1. \( R_{jl}(s) \) is additively separable in partition \( \{P_1, \cdots, P_M\} \);

2. Distribution cost, as well as fixed cost, at location \( l \) is independent of \( z_{jk}^l \) and \( x_{jk}^l \), where \( k \in P, m \neq n \).
3. \( \eta_j^l \) are independently distributed across markets.

See appendix for the proof. By Equation (3.1), it is clear that if \( \forall i \), s.t.
\[
p_{ijl} > 0, p_{ijk} > 0, l \in P_m, k \in P_n, m \neq n,
\]
then the first condition in Theorem 2 holds. In other words, if there does not exist consumer \( i \) that shops from both store \( l \) in market \( P_m \) and store \( k \) which belongs to a different market \( P_n \) (i.e. stores in different markets do not share customers), then stores \( \{1, \cdots, L\} \) are separable in partition \( \{P_1, \cdots, P_M\} \). The second condition is automatically satisfied by the specification of distribution cost \( \psi_j D_j^l \) and fixed cost \( \alpha_j x_j^l \). The condition can be violated if, for example, the distribution center has a capacity constraint and per unit cost of distributing depends on the number of stores the distribution center serves. The third condition is satisfied by the i.i.d. assumption on the cost shock \( \eta_j^l \).

Finally, I generalize the separability conditions derived above to a dynamic game setting with Bayesian Markov perfect equilibrium. The definitions are very similar to the static case, except for two differences: 1) the expected static profit function \( \mathbb{E} \pi(s) \) becomes the expected value function \( \mathbb{E} V(s_{jt}, s_{-jt}) \) in Equation (3.4), 2) instead of the market level budget constraint in one period, strategies are separable conditional on the sequence of market level budget constraint \( \{B_{jmt}, B_{-jmt}\}_{j=1}^{\infty} \), for all \( m \). The results in Theorem 1 and Theorem 2 apply. See Appendix I for details and the proofs.

3.3.5 Separability and market division

In order to apply two-stage budgeting, a sufficient condition is that the markets are separable, as discussed in the previous section. In this section, I present the tools needed to divide the U.S. national market into smaller separable markets. This is done by applying a clustering algorithm based on the separability condition and the demand model. I first define the objective function for the clustering algorithm and then I explain the steps of the algorithm to find an optimal partition of store locations given the objective function. Each element of the partition is defined as a market. Results are postponed to Section 4.2.

The main condition that needs to hold for markets to be separable is that revenue is independent across markets. In other words, stores in two different markets do not share
customers. This condition is automatically satisfied if two stores are so far away from each other, that no consumer has both stores in its choice set. Clearly, the difficulties in dividing markets arise when two markets are next to each other and consumers living close to the border of the two markets are willing to shop from either store. In reality, two neighboring stores almost never share no customer, except in areas where population density is extremely low and the stores are very far from each other. Therefore, I define an objective function for the clustering algorithm that captures how far away the partition is from being truly separable.

Define the objective function as the following,

$$\min_{\{P_1, \ldots, P_M\}} \sum_{l=1}^{L_t} [R_l^t(s, \omega) - R_l^t(s_m, \omega_m | l \in P_m)]^2,$$  \hspace{1cm} (3.10)

where $L_t$\textsuperscript{16} is the set of potential locations in period $t$, $R_l^t(s, \omega)$ is the revenue of store $l$, which depends on the set of existing stores of both firms $s$ and the determinants $\omega$ of demand which include demographic characteristics and store characteristics. Note both $s$ and $\omega$ are vectors that contain information about the entire U.S. market. The second term $R_l^t(s_m, \omega_m | l \in P_m)$ is also store $l$'s revenue, but it is computed using information of existing store locations and demand data only in the partition $P_m$. That is, it is the revenue of store $l$ if $l$ is assigned to market $P_m$. In this case, some of the spatial interdependence between $l$ and any other store $h$ that belongs to a different market $P_n$, $n \neq m$, is not accounted for. In other words, if $l$ and $h$ share any customers, those customers are restricted to shop at only one of the two stores. Consumers in block groups that have both $l$ and $h$ in their choice set are assigned to the market that their closest store belongs to\textsuperscript{17}. If $P_m$ is truly separable from the rest of the markets, then $R_l^t(s, \omega) - R_l^t(s_m, \omega_m | l \in P_m)$ is zero for all $l \in P_m$. Therefore, the sum of squared differences between $R_l^t(s, \omega)$ and $R_l^t(s_m, \omega_m | l \in P_m)$ indicates how far off a partition is from each of its elements being truly separable, or the loss of assuming the elements are separable. The solution to Equation (3.10) finds the optimal partition that minimizes this loss.

\textsuperscript{16}I kept the $t$ subscript to differentiate $L_t$ from $L$, which is all locations including both $L_t$ the potential locations and those that have been entered up to period $t$.

\textsuperscript{17}I also tried assigning consumers to the market their most preferred store belongs to. Results are similar.
Next, I introduce the clustering algorithm that attempts to find a solution to Equation (3.10). Since it is a graph partitioning problem that is NP-hard (Fortunato and Castellano, 2009), the solution is an approximation. Although there might be other approximated solutions to this problem, results in section 4.2 indicate that the clustering algorithm does reasonably well.

Start with $M = 2$. Apply a greedy algorithm which locally minimizes the objective function to find an approximated global solution to Equation (3.10). Then increase $M$ and repeat the previous step. Stop when the stopping criterion binds. Due to the complex geographic structure of the model, greedy algorithm is more suitable than other algorithms such as spectrum algorithm that has the advantage of speed but assumes additional structure of the problem. I describe the greedy algorithm and the stopping criterion in the remainder of this section.

The greedy algorithm finds the optimal partition $\{P_1, \cdots, P_M\}$ given the objective function (3.10) and the number of clusters $M$. Figure 6 demonstrates the idea for $M = 2$. Each dot is a store. The edge between a pair of dots means that the two stores share customers. The task is to cut off a few edges such that the set of locations is divided into two markets. The broken edges are selected so that the objective function (3.10) is minimized. There are two features of the problem that are important for the set-up of the greedy algorithm. First, only stores close to the border of two markets matter. The objective function is zero for stores that have all connected neighbors in the same market. This feature leads to the fact that the algorithm focuses on stores close to the borders of markets. Second, the edges between stores are weighted. The weight is the amount of interdependence between two stores and is decided by the demand data and the objective function. The weight varies across edges, thus one cannot simply minimize the number of broken edges in a graph to find the optimal partition.

The algorithm follows four steps.

1. For a given partition of locations $\{P_1, \cdots, P_M\}^t$, find all $l$ s.t. $\exists h \in C_l$, and $h \in P_n$, but $l \in P_m$, and $m \neq n$, where $C_l$ is the set of locations $l$ is connected to.

2. Reassign each $l$ in the previous step to a partition such that (3.10) is minimized keeping the assignment of all the other stores fixed. Call the new partition $\{P_1, \cdots, P_M\}^{(t+1)}$.

3. Repeat step 1 and 2 until the algorithm converges.
4. Repeat step 1, 2, and 3 for 1000 different initial partitions \( \{P_1, \cdots, P_M\}^0 \).

One nice property of this algorithm is that the resulting partition respects contiguity. If all the connected neighbors of some store(s) belong to one partition, the store(s) itself cannot be in a different partition. I.e. If \( h \in P_m, \forall h \in C_l \), then it must be that \( l \in P_m \).

Finally, I explain the stopping criterion for picking the number of partitions \( M \). As the number of clusters increases, the incremental change of loss, i.e. the value of (3.10), also increases. The stopping criterion is chosen when the sum of incremental change of loss from \( M \) to \( M + 1 \) partitions is bigger than or equal to 1\% of revenue \( R^l(s, \omega) \) for any store \( l^{18} \). It also happens to be the point at which the incremental change of loss increases dramatically in many cases.

4 Estimation and Clustering

4.1 Demand estimation

I estimate the demand model using a maximum likelihood framework. Like Holmes (2011), I assume the discrepancy between the model and the data is measurement error and that the error follows a normal distribution. Denote the measurement error by \( \epsilon_{jl} \) and observed sales of store \( l \) by \( R^{obs}_{jl} \). Then

\[
\ln(R^{obs}_{jl}) = \ln(R_{jl}) + \epsilon_{jl},
\]

where \( \epsilon_{jl} \sim N(0, \sigma^2) \).

Results of the demand estimation are presented in Table 6. The first column shows the results of the basic specification. Spending per person is on average $47 per week, which is $2444 per year. This is close to the estimate in Holmes (2011), that is of about $2150 per year in 2007 dollars. The coefficient of population density is positive and significant. The coefficient on distance is negative and significant. I interpret the coefficients using comparative statics in Table 7. Column 2 presents a different specification with a dummy variable indicating whether a store sells both general merchandise and grocery. The results are similar.

\[\text{Sensitivity checks on the 1\% criterion are to be conducted. Future research is needed to obtain a more systematic stopping criterion.}\]
I use comparative statics to illustrate the effects of distance to shops in the choice set and population density on store sales. This exercise also demonstrates how spatial competition is generated from demand. Consider a Red store located two miles away from the median block group and a new Blue store entering the market. First, I fix the population density of the block group and compute the probabilities of consumers shopping at the Red store when the distance between the new Blue store and the block group changes. Column 1 in Table 7 reports the choice probabilities when population density equals 1. Moving up across the rows, the choice probabilities decrease as distance to the Blue store decreases. This shows that the competition between the two stores intensifies as the Blue store moves closer to the consumers. The result stays the same across columns when population density takes on different values. Row 1 reports the probabilities of consumers shopping at the Red store when population density increases and distance to Blue store stays at 2 miles. From left to right, choice probabilities decrease as population density increases. This shows that the utility of choosing the outside option increases as population density increases.

4.2 Clustering results

In this section, I present the clustering result and compare it to an alternative definition of markets in the literature, Core Business Statistical Areas (CBSA).

Since the set of potential locations $L_t$ changes every period, the partitioning of markets is conducted every period. Note the existing stores are not partitioned because they are no longer in the choice set of the firms. Consequently, the spatial interdependence between existing stores and potential stores is fully accounted for. As a result, in the empirical analysis, market definitions are different in each period. The advantage of this assumption is that it is close to how managers actually make entry decisions. According to the managers and consultants I interviewed, when firms evaluate a potential store location, they define a trade area around it. The trade area is defined as the area which demand is likely to come from and which the main competing stores including the firm’s own stores and competitors’ stores are located at. Naturally, the trade area varies across time as new stores are opened each period. Therefore, the clustering procedure can also be viewed as estimating trade areas each period. The disadvantage of doing so is that it
does not allow any market level unobservables that can be controlled for using market fixed effects.

Figure 7 is a map of the northeast, mid-atlantic and (part of) southeast United States. It is the area where store density was the highest in the third quarter of 1997. The points (squares, dots, and diamonds) are the potential store locations in this period. Neighboring stores are marked with different colors or shapes to display market divisions. The CBSAs also appear in the map, in yellow. This map compares market divisions defined by the clustering algorithm to the CBSA units. In a few cases, using CBSA as a market is not very different from the clustering of the store locations. For example, the single black dot in the very northeast is the only store in its market after clustering. In this particular case, defining market as the CBSA of the store may not be a bad idea, since there are no other store around. However, in most cases, defining a CBSA as a market can be misleading. For example, the five blue dots just southwest of the black dot are defined as in one market by clustering, while they are located in three different CBSAs. Dividing the stores into three different market would be misleading since they are very close to each other, and at least two of them are almost right at the border of two CBSAs. Consumers do not confine themselves to shop within CBSA boundaries, so it is not reasonable to define markets with this restriction. Such restriction is minimized in the division of markets by applying the clustering algorithm.

Table 8 presents a few measures demonstrating the goodness of the clustering results and compares them with those by using CBSAs as markets. The measures are computed using all \( L_t \) locations in the third quarter of 1997. There are 241 potential locations. First, the total loss by clustering as a fraction of total sales, i.e. \( \frac{\sum_{l \in r^*} |S(l) - S(l \in r^*)|}{\sum_l |S(l)|} \), is less than 0.001%. The maximum store level loss as a fraction of sales is also reasonably small, 0.5%. Finally, the total number of stores affected by clustering is 55. This shows that simply excluding these locations is not a satisfying option, since it reduces the sample size by more than 20%. On the other hand, using each CBSA as a market would lead to undesirable results. The maximum store level loss is 53.0%, about a hundred times higher than that of clustering.

27
4.3 Cost estimation

In this section, I explain the estimation of the cost function. Due to the non-stationary nature of the game, the estimation takes a different approach from the two-stage procedure proposed by BBL or POB. The approach includes two main parts. It first computes the value function using a ‘rolling window’ approximation. Then the game is solved using backwards induction and the approximated continuation values. The advantage of the approach is that it is more suitable for the counterfactual of interest which is to examine preemptive incentives, and that it is closer to how managers actually make entry decisions, as will be discussed later.

First, I describe the estimation of the parameters outside the dynamic game. Recall firm’s static profit of store $l$ is given by (3.2). The gross margin, labor cost, and land cost are estimated following Holmes (2011). $\mu_j$ is computed using information in firms’ annual reports. The average gross margin for Blue and Red firm is 0.24 and 0.22, respectively. The amount of labor and land of a store is assumed to be proportional to revenue:

$$E(R_{jt}^l) = \mu_E R_{jt}^l,$$

$$L(R_{jt}^l) = \mu_L R_{jt}^l.$$  

$\mu_E$ is calibrated using labor costs listed in firms’ annual reports. $\mu_L$ is computed using census data and county property tax data of a subset of stores. See Appendix A of Holmes (2011) for details.

The estimated parameters in the dynamic game are per unit distribution cost $\psi_j$ and fixed cost $\alpha_j$ of Blue and Red firm. Recall firms choose the optimal locations given the total number of new stores to open each period by maximizing (3.4). Applying two-stage budgeting to this problem with separable markets, firms first choose the optimal locations within each market given the total number of new stores in each market, and then optimize over market level budgets. In the estimation, for computational reasons, I condition on the observed market level budgets and use the information in firms’ entry decisions within each market only. In other words, each period, firms solve Equation (3.6) for each independent market. Without computational constraints, one can in addition solve the upper level optimization problem to get the optimal number of new stores allocated to each market.
Most of the dynamic game estimation methods in the literature follow a two-stage procedure as in BBL or POB. However, the current problem does not fit in the two-stage estimation framework, for two reasons.

First, in the two-stage framework, the state transition probabilities are estimated non-parametrically in the first stage. This requires observing each state repeatedly in the sample. In the current sample, however, no state is observed more than once. The nature of the problem is non-stationary. Because both firms are expanding in the sample period, the number of stores keep growing and new stores appear every period. Although clustering and market division that are done every period make the evaluation of potential locations focus on local nearby locations and demographics, the set of potential locations are different in each period. To convert the problem into one that fits the two-stage estimation framework, one solution would be to divide different states into groups by some similarity measure and to treat each groups of states as a single state. This is not desirable for a second reason: the state contains rich geographic information that can be important for studying preemptive incentives. Location characteristics contained in the state variable include the number of competitor’s stores, stores belonging to the same firm around them, distance to those stores, distance to the distribution center, and demographic information. They are important information for identifying preemptive incentives. Pooling them into groups may bias the results.

Second, even if the estimation can be done using the two-stage method, the counterfactuals of interest cannot. Using the transition probabilities estimated in the first stage to compute the preemption counterfactual can be problematic. In the counterfactual where preemption is removed, it will be required that players optimize the entry decision without taking into account preemption motives, i.e. firms are not fully optimizing. If the transition probabilities are estimated in the first stage non-parametrically and applied in the counterfactual, it is not guaranteed that those are still the correct transition probabilities when preemption is not allowed. Therefore, the two-stage estimation method is not suitable.

One alternative way to estimate the game is to solve for the nested fixed point as in Pakes (1986) and Rust (1987). In other words, for a fixed parameter value, the dynamic game can be solved and the optimal choice probabilities $Pr(a_{jmt}|s_t)$ can be matched to the observed entry decisions in each market and each period. Then, the above step is
repeated for many parameter values to find the estimate that maximizes the likelihood of
the observed choice. The difficulty is that instead of the single agent’s dynamic problem
in Pakes (1986) and Rust (1987), the current problem also has strategic interactions
between players, which makes the value function more difficult to compute. Next, I
describe how the value function can be approximated using a ‘rolling window’ method.

Each period, in each market \( m \), firm chooses \( B_{jm} \) locations to open new stores from
the \( L_{mt} \) potential stores. The set of potential locations \( L_{mt} \) is all the locations Blue and
Red firm entered between \( t \) and the end of the sample period, \( T \). First, the value function
can be computed by solving the game using backwards induction. Starting from the last
period, \( T \), terminal values can be computed for each possible state \( s_{mT} \). I assume the
terminal value equals \( \mathbb{E}V(s_{mT})/(1-\beta) \). \( \beta \) is set to be 0.99, i.e. the annual discount factor
is 0.95. \( \mathbb{E}V(s_{mT}) \) can be computed by solving the game for the last two years of data left
out of the sample. It allows the decision in period \( T \) to be dynamic with respect to the
8 periods ahead, instead of completely static. The implicit assumption is that firms do
not foresee any more entry or change in demographics after \( T + 2 \). This is a limitation
but no more data is available.

Second, to compute the continuation value for each of the \( \binom{L_{mt}}{B_{jm}} \) states, the value
function needs to be computed for each of the possible states in the future between \( t \) and
\( T + 2 \). This is computationally infeasible. For period \( t = 1 \), \( L_t = L = 3123 \). Assume
clustering can reduce the market size to \( 30^{19} \), the number of value functions need to be
computed is

\[
\begin{align*}
\binom{L_m}{B_{1m}} \cdot \binom{L_m - B_{1m}}{B_{2m}} \cdots \binom{L_m - \sum_{\tau=1}^{T} \sum_{j} B_{j\tau}}{B_{1mT}} \cdot \binom{L_m - \sum_{\tau=t}^{T} \sum_{j} B_{j\tau} - B_{1mT}}{B_{2mT}} \\
= \binom{L_m}{B_{1m}} \cdot \binom{B_{1m}}{B_{1mT}} \cdot \binom{B_{2m}}{B_{2mT}} \cdot \binom{B_{1m} - B_{1mT}}{B_{1m(T-1)}} \cdot \binom{B_{2m} - B_{2mT}}{B_{2m(T-1)}} \cdots
\end{align*}
\]

(4.1)

where \( B_{jm} \) is the total number of new stores that belong to firm \( j \) in market \( m \), \( j = 1, 2 \).
Assume half of the stores are Blue, the first term \( \binom{30}{15} \) is on the order of \( 10^8 \). Moreover,
the game has to be solved for each parameter value in the estimation. Therefore, an
approximation of the value function is necessary for both the firm and the econometrician.

As discussed above, changing the state variables to reduce the size of the state space

---

\(^{19}\)This is already unrealistic. It implies the locations need to be divided into 100 markets, therefore
the loss from clustering must be substantial.
is not a desirable approach. One can also reduce the state space by restricting \( j \)'s choice set to be locations entered by \( j \) only. In this case, the first term in (4.1) would become 1. However, not allowing firm \( j \) to choose from \(-j\)'s observed locations rules out some of the strategic interactions between the two firms, including preemptive incentives. As an alternative, I restrict the choice set of the firm in each period using a rolling window. For each period \( t \), instead of choosing from all observed locations between \( t \) and \( T+2 \), firms choose from those entered by either firm between \( t \) and \( t+8 \). In other words, \( j \) solves the following equation in period \( t \) for each market \( m \),

\[
V(s_{jmt}, s_{-jmt}) = \max_{a_{jmt} \in A_{mt}} \left\{ \sum_{\tau=t}^{t+8} \beta^{\tau-t} \sum_{s_{-jmt} \in A_{mt}} \mathbb{E}V(s_{jmt\tau}, s_{-jmt\tau}) \right\} \cdot (4.2)
\]

s.t.

\[
\sum_{t=1}^{L_{mt}} a_{jmt}^l \leq B_{jmt},
\]

where \( L_t \) is the set of locations entered by Blue and Red firm between \( t \) and \( t+8 \), \((1 - \beta)^1_{\{\tau=t+8\}} \) is a scaling factor for terminal period \( t+8 \). This reduces the computational burden dramatically. Clustering is now done over location \( L_t \) which is all potential locations between \( t \) and \( t+8 \). The maximum size of market, \( L_{mt} \) is 12 after taking out a few city centers.

The approximation is close to how managers actually make decisions. According to the managers I interviewed, firms usually have a fairly good idea about how many stores they are planning to open and where the potential locations are in the next two years, which corresponds to 8 periods in my model. Beyond the two year window, it is difficult for managers to foresee how many new markets they are likely to enter or where the desirable locations could be.

However, the approximation imposes two restrictions on firm’s optimization problem. First, since the optimization stops at \( t+8 \), any change of the state variable after \( t+8 \), and thus any change of the continuation value, is not taken into account. The number of possible paths that need to be evaluated becomes

\[
\left( \begin{array}{c} L_{mt} \\ B_{1mt} \end{array} \right) \cdot \left( \begin{array}{c} L_{mt} - B_{1mt} \\ B_{2mt} \end{array} \right) \cdots \left( \begin{array}{c} L_{mt} - \sum_{t+\tau=t}^{t+7} \sum_j B_{jmt\tau} \\ B_{1mt+8} \end{array} \right) \cdot \left( \begin{array}{c} L_{mt} - \sum_{t+\tau=t}^{t+7} \sum_j B_{jmt\tau} - B_{1mt+8} \\ B_{2mt+8} \end{array} \right).
\]

31
Second, the possible paths firm $j$ is optimizing over between $t$ and $t + 8$ are further restricted due to the restriction of choice sets by the rolling window. All the $L_t$ in (4.3) becomes $\bar{L}_t$, and therefore the last term in (4.3) becomes 1. The computational burden is further reduced. However, the possible paths between $t$ and $t + 8$ are restricted to include only observed locations $\bar{L}_t$, that is, any locations entered after period $t + 8$ are not considered as a potential location in period $t$. I am currently working on testing how sensitive the results are to this restriction by allowing the length of the rolling window to vary.

Finally, I do a grid search through the parameter space and use maximum likelihood to estimate the parameters $\{\psi_j, \alpha_j\}$, $j = 1, 2$. Assume the cost shock of each action follows a type I extreme value distribution. Then the choice probabilities $P(s_{jmt+1}\mid s_{jmt}, s_{-jmt})$ have a closed form solution. For a given parameter value, I solve the game in (4.2) for each market, period, and firm, and match the choice probabilities to the observed choice by:

$$
\max_{\psi_j, \alpha_j} \sum_m \sum_t \sum_j \log \left( P_r(a_{jmt}\mid s_{jmt}, s_{-jmt}, \psi_j, \alpha_j) Y(a_{jmt}) \right),
$$

where

$$
P_r(a_{jmt}\mid s_{jmt}, s_{-jmt}, \psi_j, \alpha_j) = \frac{\exp(\mathbb{E}V(s_{jmt} + a_{jmt}, s_{-jmt}))}{\sum_{a_{jmt} \in A_{jmt}} \exp(\mathbb{E}V(s_{jmt} + a_{jmt}, s_{-jmt}))},
$$

and $Y(a_{jmt})$ is the indicator of the observed choice, i.e.

$$
Y(a_{jmt}) = \begin{cases} 
1 & a_{jmt} = a_{observed} \\
0 & \text{otherwise}
\end{cases}.
$$

### 4.4 Estimation results and interpretation of distribution cost and fixed cost

Table 9 presents the estimation results of distribution cost and fixed cost by firm. The distribution cost per thousand-mile is 1.61 million dollars for Blue firm and 0.68 million dollars for Red firm. Using industry sources, Holmes (2011) estimated Blue firm’s trucking cost of distribution to be 0.8 million dollars per thousand-mile. Using his model,
he estimated the total distribution cost to be around 3.5 million dollars per thousand-mile. My estimate of 1.61 is in the interval of [0.8, 3.5], and closer to the industry source of trucking cost than Holmes’ estimate. Holmes interprets his estimate of distribution cost as economies of scale, since it measures the average cost saving of locating a store 1000 miles closer to a distribution center in a single agent’s optimization problem. The smaller economies of scale in my results is mainly due to the fact that the model takes into account interactions with Red firm. Since distribution centers are located in rural areas, moving a store closer to a distribution center implies moving away from urban markets where demand is high. This could lead to giving up profitable locations to the competitor, especially, as will be shown below, if the competitor (Red firm) has an urban advantage. Therefore, taking into account the strategic interactions with Red firm, the overall economies of scale is smaller in this model.

The estimates also indicate that per unit distribution cost is lower for Red firm than for Blue firm. However, as shown in Table 2, Red stores are on average further away from their own distribution centers than Blue stores are from theirs. For example, the average per store distribution cost in 1990 is 0.22 million dollars for Blue firm and 0.15 million dollars for Red firm. The ratio 0.22/0.15 equals 0.68 which is the ratio of average sizes of Blue stores and Red stores. Thus the average per store distribution cost conditional on the size of the store is about the same for Blue and Red firm. Moreover, the average per store distribution cost translates to about 0.5% of sales for Blue firm and 1% for Red firm. Therefore, Blue firm has a more efficient distribution system than Red firm, which is consistent with findings by Bradley et al. (2002).

The average fixed cost of operating increases by 0.43 million dollars and 0.15 million dollars per year for Blue and Red firm respectively, when population density increases from 250 (25th percentile) to 700 (50th percentile) thousands people per 5-mile radius circular area in 1990. The average fixed cost per store in 1990 is about 1.84 and 0.62 million dollars, or 4% or 5% of sales, for Blue and Red firm, respectively. Since 0.62/1.84 < 0.68, Red firm has an urban advantage relative to Blue firm. This is consistent with the store characteristics comparisons in Table 2.

Standard errors are computed using bootstrap over markets. They do not include estimation errors of demand in the first stage or clustering errors in the second stage. I am currently working on including those errors in the standard errors of the structural
estimates using a simulation method\textsuperscript{20}. See Appendix II for details of the simulation method.

5 Counterfactual I: Preemptive entry

In this section, I conduct a counterfactual analysis to examine the impact of preemptive motives on discount retailers’ entry decisions. I remove preemptive motives using a one-period deviation method and compare firms’ optimal response to that of the equilibrium outcome with preemption. In other words, what the optimal entry decisions would be if the firm did not need to preempt. I find that preemptive incentives are important to firms’ entry decisions and that they lead to an average loss of production efficiency of about 1 million dollars per store, measured by combined sum of current and future profits of the two firms.

It is hard to identify preemptive incentives because they arise in a complex dynamic setting. For preemptive motives to arise, both dynamic optimization and strategic interactions between firms have to be allowed. In such settings, for example, firm \( j \) optimizes over three sets of variables: current state \((s_{jt}, s_{-jt})\) which I refer to as ‘static competition’, \( \{s_{j\tau}\}_{\tau>t} \) which is the possible state of \( j \) and where economies of scale come from, and \( \{s_{-j\tau}\}_{\tau>t} \) which is the opponent’s possible state in the future and where preemption comes from. Moreover, as preemption is a motive for acting rather than an action, it cannot be directly observed by the econometrician. Given the complex setting in which preemption arises and its unobserved nature, it is often hard to separate it from static competition between players, incentives to optimize dynamically or unobserved market characteristics. Furthermore, the evaluation of efficiency loss requires that the game setting in the counterfactual not drastically differ from the original setting such that payoffs are comparable under the two settings. Thus solving a different game in which preemption incentives are removed would not be appropriate for the purpose of this counterfactual analysis.

Next, I introduce a formal definition of preemption and present a one-period deviation method that separates preemptive motives from static competition and leads to simple payoff comparisons. I define the preemptive incentives of firm \( j \) entering a location \( l \) to be

\textsuperscript{20}I am hoping to present results in January.
the change in $Pr(a_{jt}^l|s_t)$ in response to $Pr(a_{-jt'}^l|s_t)$ in equilibrium, where $Pr(a_{jt}^l|s_t)$ is the choice probability of firm $j$ entering location $l$ at state $s_t$ in equilibrium, and $Pr(a_{-jt'}^l|s_t)$ is the same probability for $-j$ in period $t'$, $t' > t$. In other words, preemptive incentives measure how much, in equilibrium, firm $j$'s likelihood of entering location $l$ today is impacted by its opponent’s likelihood of entering the same location in the future, holding static profits constant.

The one period deviation method attempts to measure the change in $Pr(a_{jt}^l|s_t)$ when $Pr(a_{-jt'}^l|s_t)$ is set to zero. The idea is the following: for each of Blue’s observed choices, remove those choices from Red firm’s choice set for one period. Thus Blue firm knows Red would not be allowed to enter those locations for one period. Then I investigate if Blue firm has profitable deviations by delaying entry at those locations. Specifically, I solve the following equation to compute the choice probabilities of Blue without preemption,

$$V(s_{jmt}, s_{-jmt}) = \max_{a_{jmt} \in \mathcal{A}_{mt}} \left\{ \sum_{\tau=t}^{t+8} \sum_{s_{-jmt}, s_{jmt}} \mathbb{E}V(s_{jmt}, s_{-jmt}) P(s_{jmt}|s_{jmt}+a_{jmt}, s_{-jmt}) \right\}$$

s.t.

$$\sum_{l=1}^{L_{mt}} a_{jmt}^l \leq B_{jmt},$$

where $a_{jmt+1}^{obs}$ is Blue’s observed choice in period $t$. Note the market level budget constraint is held fixed, so that Blue firm is not fully optimizing. Restricting $a_{-jmt} \neq a_{jmt+1}^{obs}$ gives Blue firm an advantage over the set of locations in $a_{-jmt}$ and the firm’s payoff is at least as high as in the original equilibrium. Thus relaxing the market budget $L_{mt}$ would only lead to higher payoff. As a result, if Blue firm’s payoff increases by solving Equation (5.1), the amount of payoff increase is a lower bound, i.e. the impact of preemptive incentives shown by the above procedure is a lower bound.

For the opponent, Red firm, the optimization problem is the same as in (5.1) with $j$ and $-j$ switched. If Blue firm’s choice probabilities stay the same, Red firm would also stay in the original equilibrium. If Blue firm deviates, Red firm is allowed to respond. Note that in this case, the market budget $\bar{L}_{mt}$ is also held fixed for Red firm. This does not bias the result. Red firm is deprived by being forced to choose from a smaller set of
locations for one period. If Blue firm were to deviate and delay entry, and Red firm were fully optimizing, it would switch away to other markets, inducing a weaker presence in the current market. This would lead to higher payoff for Blue firm. Therefore the impact of preemptive incentives measured in this experiment is a lower bound.

Results are presented in Table 10. I compute the choice probabilities of Blue firm entering the observed locations when preemptive incentives are removed for one period. For cases in which it is profitable for Blue firm to deviate from the original equilibrium, I compute the payoff increase from the original equilibrium payoff. Out of 1278 locations, there are 425 locations at which preemption is observed in this experiment. There is profitable deviation for Blue firm to delay entry when those choices are taken out of Red’s choice set for one period. For those 425 locations where preemption is observed, average choice probability decreases by 0.14 compared to the choice probabilities in the original equilibrium. Blue firm’s payoff could increase by an average of 0.86 million dollars for each delayed entry, which is about a small store’s one year profits.

Then I study loss production efficiency due to preemption. Since I do not have enough data to conduct total welfare analysis, in the current counterfactual, I compute efficiency loss from firms’ perspectives only, by examining the change of sum of expected current and future profits of both firms. Results are shown in the last two rows of Table 10. Out of the 425 locations, there are 392 locations where the expected total payoff of Blue and Red firm is higher in the counterfactual than in the original equilibrium. These are the locations at which the higher payoff of Blue firm in the counterfactual is enough to compensate the lower payoff of Red firm due to its restricted choice set in one period. The total amount of payoff increase is almost 397 million dollars. On average, the efficiency loss per location is 1.01 million dollars, which is about the annual profit of a small-to-median size store. For the reasons discussed above, it is a lower bound of the efficiency loss in consequence of preemption. The market level budgets are held fixed and firms are not fully optimizing. From the perspective of the production side, preemption results in a substantial loss of production efficiency.

Next I investigate why evidence of preemption is found at 425 locations but not at others. One reason could be that preemption may not always be profitable. I compare the two sets of locations in Table 11 and refer to them as preemption and non-preemption locations. First of all, distance to Blue’s distribution center is smaller for the non-
preemption stores than for the preemption stores. The opposite is true for distance to Red’s distribution centers. This is consistent with the motives of preemption in the current experiment. Since preemption is more profitable at locations where the opponent is more likely to enter in the future, those locations are likely to be closer to Red distribution centers. On the other hand, delaying entry at these locations would mean giving up current profits, hence it is more likely to observe preemption in the current experiment at locations where current profits are low. This is consistent with longer distance to Blue’s distribution center. Therefore, in the current one-period deviation experiment, it is more likely to observe preemption at locations that are closer to Red distribution center and further away from Blue distribution center. Moreover, preemption stores also tend to locate in areas with higher store density and higher population density. This is also consistent with the fact that those are the areas that Red store is more likely to enter.

Finally, I compute the one-period deviation for Red firm. In this experiment, Blue firm is assumed to stay away from its observed choices for one period and Red firm is given the opportunity of choosing from those locations in the next period. I examine if Red firm would choose to enter those locations, and if so how much its payoff would increase. In this case, Red firm’s budget constraint for each market is allowed to adjust, since in the original equilibrium Red firm may not have entered any location in the same market. However, this is not computationally burdensome since each market is evaluated independently in this experiment. For example, when evaluating Red firm’s deviation in market $m$, all the other markets $-m$ are held fixed at the original equilibrium. Thus, to determine if it is profitable, the payoff of deviation is compared to the median expected payoff of $-m$ markets.

The results are presented in Table 12. There is about one third of the locations where it is profitable for Red firm to enter right away, had Blue firm stayed out of those locations. The average probability of entry is 0.77. The average payoff increase is 2.99 million dollars for each location, which is about a big store’s one year profits.

In the current literature, there are two methods of identifying preemption using empirical analysis. The first method, used in Schmidt-Dengler (2006), separates preemption by solving a pre-commitment game following the theoretical work of Reinganum (1981). In the pre-commitment game, firms make the entry decision in the first period for the
following $T$ periods and commit to it. The problem of this approach is two-fold. First, it does not exclude preemption completely. It prohibits a firm from responding instantaneously to the opponent’s action, but the firm is still optimizing in period 1 taking into account the possible actions of the opponent in the future. Second, since pre-commitment games demand a different setting and lead to completely different equilibria, it is difficult to compare payoffs with the ones in the original game.

The second method is to solve a single agent’s dynamic problem, as in Igami and Yang (2014). Applying their method in the current setting, Red firm would make entry decisions assuming Blue firm would not enter any market in any period. From Blue firm’s perspective, Red firm has become a part of nature and does not respond to Blue’s actions, and therefore cannot be preempted. Thus Blue firm solves a single agent’s dynamic optimization problem. This experiment does not properly separate preemptive incentives since it also precludes Red firm from responding to Blue firm’s actions in the current period, i.e. it prohibits static competition. Setting preemptive motives apart, it is not clear why, when Blue’s entry is observed, Red firm should optimize assuming Blue firm is not entering any market.

6 Counterfactual II: Subsidy Policy after Red Firm Exits

Red firm started exiting in many markets by closing stores in 2000. It has closed more than 1000 stores in the past 15 years. The store closings have a big impact on local economies (Shoag and Veuger, 2014). Local governments have proposed to subsidize Red firm to stay or other retailers such as Blue firm to enter. For example, Buffalo, NY proposed a 400,000 dollar subsidy for Red firm to stay\textsuperscript{21}. Rolling Meadows, IL managed to subsidize Blue firm to enter after their Red store closed\textsuperscript{22}. However, lots of ex-Red retail slots remained empty years after Red firm’s exit. The example of Rockledge, FL I have mentioned. Indiana Harbor Beach, FL is another example\textsuperscript{23}. The retail space of the former Red store stayed empty for 12 years. Therefore, it seems important for

\textsuperscript{21}Source: http://www.huffingtonpost.com/2012/01/26/sears-closes-cities_n_1231326.html
\textsuperscript{22}Source: goodjobfirst.org.
\textsuperscript{23}Source: http://www.floridatoday.com/story/money/business/2014/07/27/kmart-goes-next/13197001/
policy makers to better understand the impact of government subsidies on retailer’s entry decisions. The current entry model can be used to answer this question. It is also important to consider the welfare loss due to the closing of stores. Although I do not have enough data to conduct total welfare analysis, I compute the loss of consumer drive time due to store closings and compare it to the size of the observed subsidies. This section attempts to answer those two questions by computing payoff differences between ex-Red locations and the rest of potential Blue firm’s locations and deriving welfare loss to consumers caused by higher travel cost of shopping.

First, I compute the expected payoff of Blue firm entering each ex-Red locations and compare it to the expected payoff of Blue firm entering each of the other potential locations for each period between year 2000 and 2003. There are 96 Red store closings in those four years and the number of potential locations is 815. I refer to the difference between the payoff of the median store in the two groups as the ‘subsidy’, since it is the amount of extra payoff needed for the ex-Red locations to become as profitable as the other potential locations. Then I compute subsidies separately for two sub-periods: 2000-2001, and 2002-2003. The difference between those two sub-periods is that between 2000 and 2001, Red firm was still expanding, but beginning in 2002 the expansion stopped. This makes a difference for Blue firm’s incentives to enter ex-Red locations. In the first sub-period, it is clear that Blue firm knew that Red firm would not re-enter at ex-Red locations after the store closings, which removes the preemptive motives for Blue to enter those locations, compared with other potential locations. On the other hand, in the second sub-period, there is no preemptive motive for Blue firm at any potential location including the ex-Red ones since the expansion of Red firm has stopped. Thus it does not make the ex-Red locations less favorable, as in the first sub-period.

Results are reported in row 1 of Table 13. The median size of subsidies in the period between 2000 and 2003 is 3.21 million dollars per store. The same measure is higher for period 2000 and 2001 when Red firm is still expanding: 5.13 million dollars per store. The stop of Red firm’s expansion makes ex-Red locations less unattractive compared to other potential locations, with a median subsidy amount of 1.29 million dollars per store. The difference between the subsidies needed for Blue to enter in the two sub-periods gives one explanation to why some retail slots remained empty for years after a store closing. The preemptive motives lead to the unattractiveness of those locations compared to other
potential locations.

To better interpret the sizes of the subsidies, I compare them to the size of observed subsidies given to Blue firm between 2000 and 2014. The subsidy data is obtained from goodjobsfirst.org. Although the list of subsidies is incomplete and some of the subsidy sizes are approximated, it gives a general sense of the size of the subsidies. The average size is about 0.5 million dollars. Accordingly, I count the number of ex-Red locations whose payoffs are less than half a million dollar lower than the payoff of the median potential location that has never been entered by either firm. Row 2 of Table 13 reports the results. On average, there are 5.5 ex-Red locations per period that Blue firm would enter with a subsidy of 0.5 million dollars. The number drops to 3.5 for the period of 2000-2001, while an increase of 4 locations period is observed for the period of 2002-2003. Overall, the observed average subsidy does not have a big impact on Blue firm’s entry at locations where Red firm exited.

Lastly, I examine the welfare loss of consumers due to increased travel time to discount shops when Red stores close. This is not expected to measure the total welfare change due to store closings. Other factors such as employment, impact on small businesses, local government income are also affected by exits of big box retail stores (Basker, 2007, Jia, 2008). However, this analysis still helps to get a sense of whether, in general, the welfare loss is comparable to the size of subsidies. I use the demand model to compute the change of distance between a consumer and a store in the consumer’s choice set due to each of Red store’s closings. Note some consumers may switch to the outside option after a Red store closes. For those consumers, I assume the distance traveled to the outside option to be the average distance travelled by a consumer to a discount retail store, 15 miles. This measure comes from the industry survey data collected by Fox et al. (2004).

Table 14 reports the results. The average travel distance per person increases by 4.05 miles between 2000 and 2003, while the total distance increases by 870 thousands miles. The total welfare loss per year is computed using the following formula: total distance/40mph×7.25(federal min. wage)×10 trips×2(round trip)/2.5(avg. household size). Total distance divided by driving speed of 40 mile per hour is the total time of travel which is multiplied by the federal minimum wage to get the dollar value. I assume a consumer makes 10 trips per year to a discount store. Given the estimated annual spending is $2444, it seems reasonable to assume she spends about $240 on each trip.
10 trips per year is also much lower than the estimate by Fox et al. (2004) of visiting a discount store every two weeks, which squares with the goal of finding a lower bound for consumer welfare loss. Then the number is multiplied by 2 to account for round trips and divided by the average household size from census data, assuming one person shops for each household. The total welfare loss amounts to 1.26 million dollars. Although it is much smaller than the average subsidy size of 3.21 million dollars, the break down of 1.10 million dollars for period 2002-2003 is very close to the 1.29 million dollar subsidy for this period. Therefore, welfare loss to consumers due to store closings can be substantial and can make the subsidy on entry worthwhile.

7 Conclusion

This paper studies how multi-store retail chains make entry decisions, with a special emphasis on the impact of preemptive incentives.

The study is carried out in a dynamic duopoly model in which firms make entry decisions at spatially interdependent locations. It is shown that the model can be made tractable by applying two-stage budgeting and separability. Instead of using census geographic units, market divisions are inferred using machine learning tools built on separability conditions, so that the spatial interdependence across store locations is preserved. The estimation is carried out by solving the game using backwards induction and applying a ‘rolling window’ approximation to compute the value function. This model and this estimation method can be applied to other retail industries or sectors in which network effects or cost sharing are present. More generally, the application of machine learning tools in structural estimation and its impact on inference is an interesting direction for future research.

Counterfactual analyses are also conducted. The results suggest that preemptive incentives are important in multi-store retailers entry decisions and that they can lead to substantial efficiency loss. When a retailer exits a market, as frequently observed in the recent crisis, the store location becomes less attractive to other retailers due to the absence of preemptive incentives. In these cases, although consumer welfare loss from the store closings can be significant, standard government subsidies prove insufficient to encourage entry. The framework presented here can be used to assess other public policy
issues that arise in those industries the model can be applied to.

APPENDIX I: PROOFS OF THEOREMS

Proof of Theorem 1: Suppose \( \exists \sigma_j' \neq \sigma_j^* \) s.t. \( \pi(\sigma_j'(s), \sigma_{-j}^*(s), s) > \pi(\sigma_j^*(s), \sigma_{-j}^*(s), s) \). Since \( \sigma_{-j} \) and \( \sigma_j^* \) are separable, \( \exists m \in \{1, \ldots, M\} \), s.t. \( \sigma_{jm}^*(s_m, \sigma_{-jm}^*(s_m), B_m^*) \neq \sigma_{jm}^*(s_m, \sigma_{-jm}^*(s_m), B) \), and \( \sum_{l \in P_m} \pi_{jl}(\sigma_{jm}^*(s_m), \sigma_{-jm}^*(s_m), s) > \sum_{l \in P_m} \pi_{jl}(\sigma_{jm}^*(s_m), \sigma_{-jm}^*(s_m), s_m) \). But \( \sigma_{jm}^* \) is the best response of \( \sigma_{-jm}^* \), and \( \Delta \pi(s_j, s_{-j}, l)/\Delta \pi(s_j, s_{-j}, h) \) does not depend on \( (s_{jn}, s_{-jn}) \), \( \forall l, h \in P_m \) and \( m \neq n \). Thus there’s no profitable deviation by including \( s_{jn}, \forall n \neq m \), i.e. \( \sum_{l \in P_m} \pi_{jl}(\sigma_{jm}^*(s), \sigma_{-jm}^*(s_m), s) \leq \sum_{l \in P_m} \pi_{jl}(\sigma_{jm}^*(s), \sigma_{-jm}^*(s_m), s_m) \).

Proof of Theorem 2: If all conditions are satisfied, \( \pi(s) \) is additively separably in \( \{P_1, \ldots, P_M\} \). By results in Gorman (1959), \( \{1, \ldots, L\} \) is separably in \( \{P_1, \ldots, P_M\} \).

Definition A.1 Locations \( \{1, \ldots, L\} \) are separable in the partition \( \{P_1, \ldots, P_M\} \) if

\[
\frac{\Delta \mathbb{E}V(s_j, s_{-j}, l)}{\Delta \mathbb{E}V(s_j, s_{-j}, h)} = (s_j^k, s_{-j}^k) \mid \{B_{jm}, B_{-jm}\}_{j=1}^M, \forall l, h \in P_m, \forall k \notin P_m, m \neq n,
\]

where \( l, h \in P_n \), and \( k \in P_{nk} \), and

\[
\Delta \mathbb{E}V(s_j, s_{-j}, l) = \mathbb{E}V(s_j^l = 1, s_{-j}^l, s_{-j}) - \mathbb{E}V(s_j^l = 0, s_{-j}^l, s_{-j}).
\]

Definition A.2 Firm \( j \)'s strategy \( \sigma_j^* \) is separable in the partition \( \{P_1, \ldots, P_M\} \) if for given \( \sigma_{-j}, \exists \sigma_j^{1}, \ldots, \sigma_j^{M} \) s.t.

\[
\sigma_{jm}^*(s_{jm}, s_{-jm}, B_m) = \sigma_j^{s_m}(s_j, s_{-j}, B),
\]

where \( \sigma_j^* = (\sigma_j^{1}, \ldots, \sigma_j^{M}) \), \( B_m = (B_{jm}^*, B_{-jm}) \), \( B_{jm}^* = \sum_{l \in P_m} a_{jm}^l \), \( B_{-jm} = \sum_{l \in P_m} a_{-jm}^l \), \( \forall m = 1, \ldots, M \), and \( \sum_{m=1}^M B_m = B \).

Theorem A.1 If locations \( \{1, \ldots, L\} \) are separable in the partition \( \{P_1, \ldots, P_M\} \), there exists a separable equilibrium.

Proof of Theorem A.1: Results follow the proof of Theorem 1.

Theorem A.2 The location \( \{1, \ldots, L\} \) is separable in partition \( \{P_1, \ldots, P_M\} \) if the value function \( \mathbb{E}V(\cdot) \) satisfies the following conditions,

1. \( R_j^*(s) \) is additively separable in partition \( \{P_1, \ldots, P_M\} \),
2. Distribution cost and fixed cost at location \( l \) is independent of \( z_j^k \) and \( x_j^k \), where \( k \in P_n, m \neq n \),
3. \( \eta_j^l \), are independently distributed across markets.
Proof of Theorem A.2: Prove by induction. Rewrite the value function as

$$EV(s_{ht}, s_{-jt}) = \sum_{\tau = t}^{\infty} \beta^{\tau-t} \left( \sum_{s_{\tau}} \mathbb{E}\pi(s_{\tau})P(s_{\tau}|s_{t}) \right) = \sum_{\tau = t}^{\infty} \beta^{\tau-t}EV_{\tau}(s_{t}).$$

The first term in the outer sum, $EV_{\tau}(s_{t}) = \mathbb{E}\pi(s_{t})$ when $\tau = t$, is separable in $\{P_{1}, \ldots, P_{M}\}$ by Theorem 2. Assume $EV_{\tau}(s_{t})$ is separable in $\{P_{1}, \ldots, P_{M}\}$ for $\tau = T$, then $\sum_{s_{\tau}} \mathbb{E}\pi(s_{\tau})P(sT|s_{t})$ is separable. Apply two-stage budgeting, $P(s_{T}^{t}|s_{t}, \{B_{mt}\}_{t=1}^{T})$ does not depend on $(s_{jt}^{k}, s_{-jt}^{k})$, $\forall l \in B_{m}, k \in B_{n}, m \neq n$. It is left to show $EV_{T+1}(s_{t})$ is separable.

$$EV_{T+1}(s_{t}) = \sum_{s_{T+1}} \mathbb{E}\pi(s_{T+1})P(s_{T+1}|s_{t})$$

$$= \sum_{s_{T+1}} \sum_{s_{T}} \mathbb{E}\pi(s_{T+1})P(s_{T+1}|s_{T})P(s_{T}|s_{t}).$$

Then

$$\Delta EV_{T+1}(s_{jt}, s_{-jt}, l_{T+1})$$

$$= \sum_{s_{T}} \sum_{s_{T+1}} \left[ \mathbb{E}\pi(s_{T} + l_{T+1}, s_{-jT+1}) - \mathbb{E}\pi(s_{jt}, s_{-jT}) \right] P(s_{-jT+1}|s_{T})P(s_{T}|s_{t}),$$

where $l_{T+1}$ indicates new store $l$ opened in period $T + 1$, and $l \in P_{m}$. $\mathbb{E}\pi(s_{T} + l_{T+1}, s_{-jT+1})$ and $\mathbb{E}\pi(s_{jt}, s_{-jT})$ are separable by the separability of the static profit. As a result,

$$\mathbb{E}\pi(s_{T} + l_{T+1}, s_{-jT+1}) - \mathbb{E}\pi(s_{jt}, s_{-jT}) = \mathbb{E}\pi(s_{jmT} + l_{T+1}, s_{-jmT+1}) - \mathbb{E}\pi(s_{jmT}, s_{-jmT}),$$

where $s_{jmT} = \{s_{jt}^{h} | h \in P_{m}\}$, and $s_{-jmT} = \{s_{-jt}^{h} | h \in P_{m}\}$. Note

$$P(s_{-jT+1}|s_{T}, B_{mT+1})$$

$$= P \left[ \mathbb{E}\pi(s_{jmT+1}, s_{-jmT+1}) - \mathbb{E}\pi(s_{jmT+1}, s_{-jt}) \geq \max_{s'_{jmT+1}} (\mathbb{E}\pi(s_{jmT+1}, s'_{jmT+1}) - \mathbb{E}\pi(s_{jmT+1}, s_{-jmT})) \right],$$

where $s_{jmT+1} = s_{jt} + l_{t+1}$, and $s'_{jmT+1} \in \{s_{jt+1}|s_{jt} = s_{jt} + h_{T+1}, h \in P_{m}\}$,

$$= P \left[ \mathbb{E}\pi(s_{jmT+1}, s_{-jmT}) - \mathbb{E}\pi(s_{jt+1}, s_{-jt}) \geq \max_{s'_{jmT+1}} (\mathbb{E}\pi(s_{jmT+1}, s_{-jmT+1}) - \mathbb{E}\pi(s_{jmT+1}, s'_{jmT})) \right],$$

where $s_{jmT+1} = s_{jt+1}l_{t+1}$. Thus $\sum_{s_{jt+1}} [\mathbb{E}\pi(s_{jt} + l_{T+1}, s_{jt+1}) - \mathbb{E}\pi(s_{jt}, s_{jt})] P(s_{jt+1}|s_{jt})$ is additively separable in $\{P_{1}, \ldots, P_{M}\}$. Since $EV_{T}(s_{t})$ is separable,

$$\frac{\Delta EV_{T+1}(s_{jt}, s_{-jt}, l_{T+1})}{\Delta EV_{T+1}(s_{jt}, s_{-jt}, h_{T+1})} \perp (s_{jt}^{k}, s_{-jt}^{k}) \{\{B_{jmT}, B_{jmT}^{T+1}\}_{t=1}^{T+1}\},$$

$\forall l, h \in P_{m},$ and $k \in P_{n}, m \neq n.$

Appendix II: Simulation method for computing standard errors

In this section, I describe how the first stage estimation error and second stage clustering error can be accounted for in the standard errors of the structural estimate in the third stage. It is a simulation method and has four steps.
1. Denote $\hat{\beta}$ and $f(\hat{\beta})$ the estimated demand parameter and its distribution from the first stage estimation. Take $R$ draws of $\hat{\beta}$, $\{\hat{\beta}^r\}_{r=1}^R$, from $f(\hat{\beta})$.

2. Recompute market divisions using the clustering algorithm described in Section 3.3.5 for a given $\hat{\beta}^r$. Denote the market divisions by $\{P^r_1, \ldots, P^r_M\}$.

3. Given demand estimate $\hat{\beta}^r$ and market division $\{P^r_1, \ldots, P^r_M\}$, estimate the structural model to get $(\hat{\psi}^r, \hat{\alpha}^r)$ and its distribution $f(\hat{\psi}^r, \hat{\alpha}^r)$.

4. Repeat the previous two steps $R$ times and compare $f(\hat{\psi}^r, \hat{\alpha}^r)$ to see if the first stage and second stage errors have an impact on the standard errors of $(\hat{\psi}^r, \hat{\alpha}^r)$.

Note the clustering error in the second stage is treated as a machine error. To properly account for the clustering error, one would need model store locations on random field. This is an interesting direction for future research.
### Tables and Figures

Table 1: Comparison of Blue firm and Red firm in 2001.

<table>
<thead>
<tr>
<th>Stores and distribution centers in 2001</th>
<th>Blue</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of stores</td>
<td>2698</td>
<td>1883</td>
</tr>
<tr>
<td>Total number of distribution centers</td>
<td>35</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 2: Location comparisons between two firms in sample period 1985-2001: median store characteristics measured in 2001.

<table>
<thead>
<tr>
<th></th>
<th>Blue</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance to closest competitor’s store</td>
<td>8.38</td>
<td>3.46</td>
</tr>
<tr>
<td>std. dev.</td>
<td>20.11</td>
<td>11.22</td>
</tr>
<tr>
<td>distance to closest same firm’s store</td>
<td>11.90</td>
<td>10.16</td>
</tr>
<tr>
<td>std. dev.</td>
<td>15.12</td>
<td>20.79</td>
</tr>
<tr>
<td>total number of stores</td>
<td>1983</td>
<td>1140</td>
</tr>
<tr>
<td>number of same firm’s stores in 30mi</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>number of any stores in 30mi</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>population density (10^5)</td>
<td>1.04</td>
<td>8.16</td>
</tr>
<tr>
<td>std. dev.(10^5)</td>
<td>3.20</td>
<td>1.86</td>
</tr>
<tr>
<td>distance to distribution center</td>
<td>98.47</td>
<td>126.56</td>
</tr>
<tr>
<td>std. dev.</td>
<td>71.41</td>
<td>122.02</td>
</tr>
<tr>
<td>total number of distribution centers</td>
<td>35</td>
<td>18</td>
</tr>
</tbody>
</table>
Table 3: Summary statistics of block group demographics.

<table>
<thead>
<tr>
<th></th>
<th>1980</th>
<th>1990</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean population</td>
<td>0.83</td>
<td>1.11</td>
<td>1.35</td>
</tr>
<tr>
<td>mean income per capita</td>
<td>14.73</td>
<td>18.56</td>
<td>21.27</td>
</tr>
<tr>
<td>mean share African-American</td>
<td>0.10</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>mean share elderly</td>
<td>0.12</td>
<td>0.14</td>
<td>0.13</td>
</tr>
<tr>
<td>mean share young</td>
<td>0.35</td>
<td>0.31</td>
<td>0.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>269,738</th>
<th>222,764</th>
<th>206,960</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of observations</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Evidence of preemptive entry: Control variables

Blue firm’s own store network and store density:
- distance to the closest distribution center
- distance to the closest Blue store
- number of Blue stores within 30 and 50 miles

Competitor’s store network and store density:
- distance to the closest Red firm’s distribution center
- distance to the closest Red firm’s store
- number of Red stores within 30 and 50 miles

Location characteristics:
- local wage and rent at time of opening
- local population and demographics within 30 miles of the location
Table 5: Evidence of preemptive entry: Blue firm’s timing of store openings

<table>
<thead>
<tr>
<th>Duration before store opening (Q), 1985-2001</th>
<th>Cox Hazard Model</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>distance to closest red distribution center</td>
<td>-0.016</td>
<td>1.179</td>
</tr>
<tr>
<td>(0.008)</td>
<td>(0.246)</td>
<td></td>
</tr>
<tr>
<td>tot no. stores around red distribution center</td>
<td>0.002</td>
<td>-0.021</td>
</tr>
<tr>
<td>(0.001)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>distance to closest blue distribution center</td>
<td>-0.011</td>
<td>-0.470</td>
</tr>
<tr>
<td>(0.002)</td>
<td>(0.435)</td>
<td></td>
</tr>
<tr>
<td>distance to closest blue store</td>
<td>0.533</td>
<td>-2.424</td>
</tr>
<tr>
<td>(0.133)</td>
<td>(1.132)</td>
<td></td>
</tr>
<tr>
<td>no. of blue stores within 30mi</td>
<td>-0.048</td>
<td>0.694</td>
</tr>
<tr>
<td>(0.016)</td>
<td>(0.171)</td>
<td></td>
</tr>
<tr>
<td>no. of blue stores within 50mi</td>
<td>-0.129</td>
<td>1.450</td>
</tr>
<tr>
<td>(0.010)</td>
<td>(0.100)</td>
<td></td>
</tr>
<tr>
<td>distance to closest red store</td>
<td>0.632</td>
<td>-6.224</td>
</tr>
<tr>
<td>(0.120)</td>
<td>(1.691)</td>
<td></td>
</tr>
<tr>
<td>no. of red stores within 30mi</td>
<td>0.041</td>
<td>-0.360</td>
</tr>
<tr>
<td>(0.016)</td>
<td>(0.154)</td>
<td></td>
</tr>
<tr>
<td>no. of red stores within 50mi</td>
<td>-0.068</td>
<td>0.682</td>
</tr>
<tr>
<td>(0.012)</td>
<td>(0.119)</td>
<td></td>
</tr>
<tr>
<td>local rent</td>
<td>0.000</td>
<td>0.002</td>
</tr>
<tr>
<td>(0.000)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>local wage</td>
<td>-0.078</td>
<td>0.644</td>
</tr>
<tr>
<td>(0.016)</td>
<td>(0.150)</td>
<td></td>
</tr>
<tr>
<td>no. of blue stores by 2002</td>
<td>0.113</td>
<td>-1.278</td>
</tr>
<tr>
<td>(0.008)</td>
<td>(0.078)</td>
<td></td>
</tr>
<tr>
<td>no. of red stores by 2002</td>
<td>0.023</td>
<td>-0.162</td>
</tr>
<tr>
<td>(0.010)</td>
<td>(0.095)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>61544</td>
<td>1983</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>0.66</td>
</tr>
</tbody>
</table>
Table 6: Demand estimates

Average weekly sales in $1000s by store

<table>
<thead>
<tr>
<th></th>
<th>specification1</th>
<th>specification2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma^2 )</td>
<td>0.082</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>spending per person</td>
<td>0.047</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>grocery dummy</td>
<td></td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000)</td>
</tr>
<tr>
<td>constant</td>
<td>-3.149</td>
<td>-3.151</td>
</tr>
<tr>
<td></td>
<td>(0.253)</td>
<td>(0.269)</td>
</tr>
<tr>
<td>popden</td>
<td>1.270</td>
<td>1.270</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>popden^2</td>
<td>-0.022</td>
<td>-0.020</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>per capita income</td>
<td>0.009</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>black</td>
<td>0.246</td>
<td>0.318</td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>old</td>
<td>-0.522</td>
<td>-0.566</td>
</tr>
<tr>
<td></td>
<td>(0.279)</td>
<td>0.285</td>
</tr>
<tr>
<td>young</td>
<td>0.052</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>(0.336)</td>
<td>(0.344)</td>
</tr>
<tr>
<td>size</td>
<td>0.504</td>
<td>0.505</td>
</tr>
<tr>
<td></td>
<td>(0.198)</td>
<td>0.200</td>
</tr>
<tr>
<td>blue</td>
<td>0.312</td>
<td>0.312</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>new</td>
<td>-0.117</td>
<td>-0.107</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>distance</td>
<td>-0.440</td>
<td>-0.441</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>distance * popden</td>
<td>0.020</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.843</td>
<td>0.845</td>
</tr>
</tbody>
</table>

Number of stores=4750
Number of blockgroups=202020
Standard errors are in parenthesis.

Table 7: Demand comparative statics

Probabilities of shopping at Red store

<table>
<thead>
<tr>
<th>Distance</th>
<th>1</th>
<th>10</th>
<th>50</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.18</td>
<td>0.14</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>0.34</td>
<td>0.21</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>10</td>
<td>0.80</td>
<td>0.33</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>20</td>
<td>0.88</td>
<td>0.34</td>
<td>0.08</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Table 8: Clustering results comparison with CBSA markets

<table>
<thead>
<tr>
<th></th>
<th>Clustering</th>
<th>CBSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max loss per location</td>
<td>0.005</td>
<td>0.530</td>
</tr>
<tr>
<td>( \max_l \left</td>
<td>S(l) - S(l \in r^*) \right</td>
<td>/ S(l) )</td>
</tr>
<tr>
<td>Total loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \frac{\sum_l</td>
<td>S(l) - S(l \in r^*)</td>
<td>}{\sum_l S(l)} )</td>
</tr>
<tr>
<td>Affected locations</td>
<td>(\sum_l \mathbb{1}_{S(l) \neq S(l \in r^*)})</td>
<td>55</td>
</tr>
<tr>
<td>Total locations</td>
<td>241</td>
<td>241</td>
</tr>
</tbody>
</table>

Table 9: Distribution and fixed cost estimates

<table>
<thead>
<tr>
<th>Parameter estimates and 95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue firm’s distribution cost ($1000/\text{mi}$)</td>
</tr>
<tr>
<td>Red firm’s distribution cost ($1000/\text{mi}$)</td>
</tr>
<tr>
<td>Blue firm’s fixed cost ($\text{M}$)</td>
</tr>
<tr>
<td>Red firm’s fixed cost ($\text{M}$)</td>
</tr>
<tr>
<td>Number of observations</td>
</tr>
<tr>
<td>likelihood ratio index</td>
</tr>
</tbody>
</table>

s.e. are computed using bootstrap and does not include the errors from first stage demand estimation or second stage clustering.

Table 10: Preemption: one period deviation of Blue firm

<table>
<thead>
<tr>
<th>Choice probabilities and payoffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average payoff increase ($\text{M}$)</td>
</tr>
<tr>
<td>Average choice probability decrease</td>
</tr>
<tr>
<td>Number of delayed entries</td>
</tr>
<tr>
<td>Total number of obs.</td>
</tr>
<tr>
<td>Number of Blue stores out of 425</td>
</tr>
<tr>
<td>Efficiency loss ($\text{M}$)</td>
</tr>
</tbody>
</table>
Table 11: Preemption vs. no preemption: location comparison

<table>
<thead>
<tr>
<th>Characteristics of preemption and no preemption locations</th>
<th>Preemption</th>
<th>No preemption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to Blue DC (mi)</td>
<td>217.94</td>
<td>208.50</td>
</tr>
<tr>
<td>Distance to Red DC (mi)</td>
<td>273.22</td>
<td>287.35</td>
</tr>
<tr>
<td>Total Store density</td>
<td>24.62</td>
<td>22.56</td>
</tr>
<tr>
<td>Blue store density</td>
<td>12.14</td>
<td>11.55</td>
</tr>
<tr>
<td>Population density (1000)</td>
<td>175.48</td>
<td>172.12</td>
</tr>
<tr>
<td>Number of observations</td>
<td>425</td>
<td>853</td>
</tr>
</tbody>
</table>

Table 12: Preemption: response of Red firm if Blue did not enter

<table>
<thead>
<tr>
<th>Change of choice probabilities and payoff</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average payoff increase ($M)</td>
<td>2.99</td>
</tr>
<tr>
<td>Probability of entry</td>
<td>0.77</td>
</tr>
<tr>
<td>Number of entries</td>
<td>472</td>
</tr>
<tr>
<td>Total number of obs.</td>
<td>1278</td>
</tr>
</tbody>
</table>

Table 13: Subsidies before and after Red firm stops expanding

<table>
<thead>
<tr>
<th>Average subsidies per store</th>
<th>Total 2000-2001</th>
<th>2002-2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median predicted subsidy ($M)</td>
<td>3.21</td>
<td>5.13</td>
</tr>
<tr>
<td>Number of subsidies ≤ 0.5M per period</td>
<td>5.5</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Table 14: Consumer welfare loss due to store closings

<table>
<thead>
<tr>
<th>Consumer drive time loss per store per year</th>
<th>Total 2000-2001</th>
<th>2002-2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance per person (mi)</td>
<td>4.05</td>
<td>4.10</td>
</tr>
<tr>
<td>Total distance (10^5mi)</td>
<td>8.70</td>
<td>9.80</td>
</tr>
<tr>
<td>Total welfare loss ($M)</td>
<td>1.26</td>
<td>1.42</td>
</tr>
</tbody>
</table>
Figure 1: Blue stores and distribution centers in 2001

Figure 2: Red stores and distribution centers in 2001
Figure 3: Blue store openings by year 1985-2001

Figure 4: Red store openings by year 1985-2001
Figure 5: Book value of total assets, 1985-2002

Figure 6: Graph partitioning\textsuperscript{a}

\textsuperscript{a}Fortunato and Castellano, 2009
Figure 7: Markets by clustering and CBSAs, 1997Q3
References


