Spatial Competition and Preemptive Entry in the Discount Retail Industry*

Fanyin Zheng†

February 2016

Abstract

Big box retail stores have a 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 store location decisions of those firms, examine the role of preemptive incentives, and evaluate the impact of government subsidies on those decisions. I model firms' store location decisions using a dynamic entry game. It extends the empirical models of dynamic oligopoly entry by allowing for spatially interdependent entry and introducing machine learning tools to infer market divisions from data. 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 loss of producer surplus due to preemption is about one million dollars per store.

JEL Classification: L13, C81, L81
Keywords: Dynamic games, Entry, Clustering, Preemption, Chain store

*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 seminar participants at Harvard University, Boston College, Columbia University, NYU, Penn State, TSE, UCLA, University of Minnesota, University of Wisconsin - Madison, for helpful comments and suggestions. Data purchased for this research is generously supported by the LEAP at Harvard University and Harvard Business School.

†Columbia Business School, 412 Uris Hall, 3022 Broadway, New York, NY 10027. fanyin.zheng@columbia.edu.

1 Introduction

The discount retail industry has been a fast growing sector of the U.S. economy. It first started to develop in the 1960s and is generating revenue of over a hundred billion dollars per year today (U.S. Census). The total revenue of Walmart, for example, is three percent of U.S. GDP (Walmart Annual Report). Such fast growth has had a large impact on local economies, including consumer welfare, employment, and small businesses (Basker, 2007; Hausman and Leibtag, 2007; Jia, 2008; Neumark *et al.*, 2008). Multi-store retail chains also receive large subsidies from local governments (Shoag and Veuger, 2014). Thus, whether subsidies affect discount retailers' entry decisions or, more generally, how the firms make entry decisions becomes 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 study the store location decisions of two discount retailers that are among the largest in the country. Due to the proprietary nature of the data, the identity of the two firms will not be disclosed. Instead, I will refer to them as Blue firm and Red firm. The two firms are the closest competitors in the industry and have both experienced long periods of fast growth.[1] Blue firm succeeded in competing against Red firm because it carefully chose store locations, exploited economies of density and, possibly most interesting, 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 otherwise enter. As a consequence, the second and more specific goal of this paper is to investigate preemptive incentives in multi-store retailers' entry decisions, to quantify their impact on those entry decisions, and to assess the loss of producer surplus due to preemption.

The definition of preemptive entry in this paper 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. There has been a large theoretical literature studying preemption games, but the empirical work has lagged behind. There is little evidence that preemptive incen-

[1] 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.

tives are important to decision makers, largely because of the difficulty of dynamic game estimation. In this study, I introduce various empirical methods to estimate a model of dynamic location game and provide both descriptive and model-based evidence of preemption.

The model has two main features. First, it is a dynamic duopoly model that allows for strategic interactions between firms. This is necessary because preemptive incentives cannot be studied in either a dynamic single-agent or a static game setting. Second, in the model, stores are spatially interdependent through demand and firm-level entry decisions. This feature fits the nature of the discount retail industry, in which firms operate multiple stores that often locate close to each other. Moreover, it generalizes the existing entry models by allowing for interdependent entry decisions in multiple markets. The model can be applied to the study of many other industries in which firms produce multiple products, run multiple plans, or operate in multiple locations.

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. The methods developed in the paper to solve this problem are sketched below. These methods are particularly interesting because they are applicable in estimating many other dynamic oligopoly games - in particular, those with out-of-sample counterfactual predictions.

First, I apply two-stage budgeting and separability conditions to decentralize firms' entry decisions across markets. I show that, conditional on optimal market-level budget, if markets are separable, entry decisions are 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 the demand data, and apply it to defining market divisions, while preserving the spatial interdependence across stores within each market. To the best of my knowledge, it is the first paper to apply machine learning tools to infer market division based on observed consumer demand. The method is applicable to many settings in which market definition affects the subsequent analysis of firm behavior and consumer welfare, and it also can be crucial to antitrust analysis.

Finally, I employ a 'rolling window' approximation to compute value functions. That is, instead of optimizing over an infinite horizon, I assume that firms optimize over a fixed number of periods and approximate the continuation value using scaled terminal values. This restricts the set of potential paths of choices over which each firm is optimizing, but the approximation is consistent with how managers actually make decisions. Due to the non-stationary nature of the problem, the dynamic game cannot be estimated in a two-stage procedure, as in Bajari *et al.* (2007) (BBL) or Pakes *et al.* (2007) (POB). 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 evaluate subsidy policies. The first counterfactual analysis quantifies preemptive incentives by removing them (partially) from one firm's optimization problem and comparing the result to the original equilibrium. The challenge is that preemption is a motive instead of an action. Thus, it is difficult to distinguish it from other optimization motives in the entry decision. The proposed solution is to use a one-period deviation approach to identify preemption. If preemptive incentives exist, when they are removed, firms would deviate in the counterfactual from the original equilibrium choice. This provides a lower bound to the size of preemptive incentives. The results show that, on average, preemption costs Blue firm 0.86 million dollars per store, which is equivalent to a small store's one-year profits. When preemption is removed, the combined current and future profits of the two firms increase by 397 million dollars, which is about one million dollars per store. Thus, the findings suggest that preemptive incentives are important in multi-store retailers' entry decisions and that preemptive entry can lead to a substantial loss of producer surplus.

In a second counterfactual analysis, I evaluate the subsidy policies proposed by local governments to encourage Blue firm's entry 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 that Blue firm needs to enter. Finally, I compute consumer welfare loss from a 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 that Blue firm received in the past.

This paper contributes to the literature on the discount retail industry. Holmes (2011) shows the importance of economies of scale in Walmart's expansion, using a single-agent dynamic optimization model. Jia (2008) studies the impact of Walmart and Kmart on small business, by solving a static game between those two firms. Ellickson *et al.* (2013) and Zhu and Singh (2009) also study 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, which are studied in the papers mentioned above. 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 was pioneered by Bresnahan and Reiss (1991) and Berry (1992). In most of the more recent literature, such as Mazzeo (2002) and 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 for interdependence across entry decisions, but the interdependence is assumed to be positive and linear in store density. This paper, however, allows for more general forms of interdependence. Moreover, it derives the interdependence from the decisions of consumers and firms. The empirical model it proposes can be applied to many other industries in which firms operate in multiple markets, and the entry decisions across markets are interdependent.

This paper also contributes to the empirical literature on preemptive incentives, which is considerably smaller than the theoretical literature. Schmidt-Dengler (2006) studies preemptive incentives in hospitals' adoption of MRI. 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. This method allows for static strategic interactions between firms and avoids solving for a different equilibrium, which keeps payoffs comparable.

As for the theoretical tools used in the paper, the two-stage budgeting and separability results are derived using Gorman's (1971) classic theorems on consumption problems.

This paper extends the main theorems in Gorman (1959, 1971) to 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 and has potential applications in the antitrust literature.

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 and the application of two-stage budgeting and separability, and it 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. In the next 40 years, the industry saw a significant growth, with total sales of discount stores peaking at 137 billion dollars in 2001 (Census, Annual Retail Trade Survey). Much of the growth is contributed by a few companies in the industry. In 2002, the four largest firms controlled 95 percent of sales (Census, Economic Census).

The two firms that this paper studies, Blue firm and Red firm, are among the four largest and have followed the path of growth of the industry. Blue firm has had a particularly interesting pattern of expansion. It was very small at the beginning of the industry, with fewer than 300 stores in the 1980s when Red firm already had more than 1000 stores. But it surpassed Red firm in the 1990s and became one of the largest employers in the country. In Table 1, which presents the total number of stores and distribution centers of both firms 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 firm and Red firm compete largely 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 that this paper will address.

Discount retail stores are known to have a large impact on local economies, as consumers benefit from their 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. was 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 closing 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 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 various forms, including sales tax rebates, property tax rebates, 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, such as Blue firm, so that it would enter. For example, Buffalo, NY proposed a 400,000 dollar subsidy to Red firm for it to stay[2]. In some places, large amounts of retail space stayed empty for years. In Rockledge, FL, for example, the former Red store has been empty for 11 years.[3] 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.

2.2 Data

Data limitations in the discount retail industry heavily constrain 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

[2]Source: www.huffingtonpost.com/2012/01/26/sears-closes-cities_n_1231326.html
[3]Source: www.floridatoday.com/story/money/business/2014/07/27/kmart-goes-next/13197001/

distribution center locations and time of opening between 1985 and 2003: these data for Blue firm come from Holmes (2011). The corresponding data for Red firm come from three sources. First, addresses and time of store openings are from infoUSA in 2002.[4] Second, I double-check the addresses and time of opening of each store using the annual report between 1984 and 2001. This step was necessary because there were 96 Red closings after 2000, and some of the stores are missing from 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.[5] 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)[6]. 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 the contiguous United States, as a snapshot at 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 in 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 that compete in many local markets across the nation.

The sample consists of Blue and Red store openings between 1995 and 2001. In this period, 1984 Blue and 1140 Red stores opened. Store openings between 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.[7]

Table 2 provides the summary statistics of the characteristics of the sample by firm.

[4] Red firm stopped opening stores after 2002.

[5] For the 12 Red stores that I could not find information about, I assumed the time of opening to be the first quarter of the year in which it first appeared in Chain Store Guide, and the time of closing to be the first quarter of the year in which they first disappeared.

[6] Distribution centers are EPA-regulated facilities.

[7] 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.

The characteristics are measured for the median store in 2001. First, 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 do from Red. 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 are 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 firms 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 in the discount retail industry and, as a consequence, have been used by other researchers (Ellickson *et al.*, 2013; Holmes, 2011). For stores that sell both general merchandise and groceries, 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 from the 1980, 1990, and 2000 decennial censuses. A block group is a geographic unit that has a population between 600 and 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 are constructed using average retail wage by county in the County Business Patterns between 1985 and 2003. Rent data are 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 that 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 is 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 these 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 interpreted as preemption. First, a firm can open more stores than otherwise optimal so that the opponent cannot enter the same market. Second, a firm can cluster its stores–i.e., have higher store density than otherwise optimal - to deter the competitor from entering. Finally, a firm can open a store earlier than otherwise optimal, so that the competitor cannot enter the same market. It is difficult to find evidence of the first two behaviors using reduced-form regression because it is hard to separate store quantity and store density from unobserved market profitability. Therefore, I focus on the third type of behavior - the timing of store opening. More precisely, I choose to study Blue firm's store opening times instead of Red firm's, for two reasons. First, Blue firm's fast growth and high store density suggests that it is more likely to have engaged in preemptive behavior, as described in the previous section. Second, during the observed time period, Blue firm has more observations of new store openings than Red firm. Thus, it is easier to find evidence of preemption if there is any.

Next, I describe how preemptive incentives can be identified in the regression analysis.

The goal is to determine whether Blue firm is more likely to enter a location earlier than otherwise if Red firm is likely to enter that location–i.e., the third type of preemptive behavior described above. The variable of interest is the time of each Blue store's opening. The difficulty is to find a location characteristic of each store that 1) affects Red firm's payoff of opening a store at that location and, therefore, Blue firm's dynamic payoff if it does not enter the location in the current period; and 2) does not directly affect Blue firm's profits from entry in the current period–i.e., is not correlated with the unobserved profitability of opening stores at the location for Blue firm. 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–i.e., the preemptive incentives in Blue firm's entry decision. Note that if such a location characteristic affects the timing of Blue firm's store opening, it implies that preemptive incentives exist. But since it is only one way that Blue firm can preempt Red firm from entering certain locations, it provides a lower bound of the size of the preemptive incentives. Since the observations in the sample are those locations that Blue firm eventually entered, the regression captures the incentives for firms to manipulate the order of store openings for strategic reasons.

One variable that satisfies these conditions is the distance between a Blue firm's store and the closest Red firm's distribution center. Consider the location of the Blue store: its distance to the Red firm's distribution center affects Red firm's cost of opening a store at that location and, therefore, the likelihood of Red firm's entry at that location. On the other hand, this distance, in general, does not directly impact Blue firm's profits from entry at the location. 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 around the location of the Blue store. To solve the problem, I include in the regression a control variable that approximates the 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.[8] The argument is that, conditional on the

[8]The underlying assumption in the analysis is that the unobserved market profitability that is correlated 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 demand did not change very much over the sample period.

total number of stores by the end of the sample period, the location of the closest Red distribution center is not correlated with the unobserved market profitability of the Blue store location of interest.

The Cox hazard model is applied to examine the impact of the distance to the closest Red distribution center on a Blue store's opening time. The dependent variable is the duration from the beginning of the sample to the time of store opening for each Blue store l, measured by 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 observation period, the distance to the Red distribution center is time-dependent. Thus, each observation is a location l observed in period t. Let h_{lt} be 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 $h_0(t)$ 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,[9] Red firm's store characteristics and other location characteristics such as wage, rent, and demographics. (See Table 4 for detailed descriptions.)

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 percent. 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 the Red distribution center 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, then Blue firm is more likely to enter the location earlier than otherwise. This is suggestive evidence of preemption.

[9] Since Blue firm was also expanding its distribution center network, the distance to Blue's distribution centers is also time-dependent.

3 Model

3.1 Overview

The model consists of two parts, a demand model and an entry model. The demand model is needed to account for revenue differentials among the existing and potential stores. It includes detailed demographic and geographic information about consumers, which allows for cannibalization across stores and, therefore, spatially interdependent store revenue.

Firms' entry decisions across locations are also spatially interdependent and, thus, modeled at the firm level instead of at the individual store or location level. A dynamic discrete-choice game framework is adopted to account for the dynamic and strategic aspects of the entry decisions. However, solving such a model proves to be computationally difficult.

To make the model tractable, a procedure similar to two-stage budgeting in consumer demand (Gorman, 1971) is applied to decentralize the entry decisions from firm level to market level. This method also mimics chain store managers' decision-making process. Based on a set of sufficient conditions for two-stage budgeting to be valid, markets are defined using a clustering algorithm. In other words, the algorithm finds the market division that minimizes the loss of store interdependence across markets. With two-stage budgeting and clustered markets, the computational burden is significantly reduced, which makes the empirical analysis possible.

Section 3.2 describes the demand model, and Section 3.3 explains the entry model.

3.2 Demand

There are two main ways to model demand based on 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 different markets. 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. The model utilizes detailed geographic information about consumers and stores, and generates spatial interdependence in revenue across store locations. The drawback of this model is that it does not allow for unobserved preference heterogeneity due to the difficulty of computing a different set of choice probabilities for each consumer.

I choose the latter approach because spatial interdependence is important for modeling chain stores' entry decisions. When evaluating the profitability of a potential store, firms need to take into account the existing stores nearby, 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, as people living in Boston or Cambridge can easily shop at any store located in either of the two cities. In other words, these stores compete for the shopping dollars of the same group of consumers.

The drawback of this approach is that it does not allow for unobserved preference heterogeneity. For example, it 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 difficult to implement. This is because there are over 200,000 block groups (i.e., unites of consumers) in the continental U.S., and each block group has a different choice set and a different set of choice probabilities. For each value of the parameters, evaluating those choice probabilities is computationally consuming. However, interaction terms in the regression and consumer-specific choice set can mediate the lack of the random coefficients 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 was 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 the log of a thousand people[10] within five miles of block group i. When population density varies, the

[10]Block groups with fewer than 1000 people are grouped together.

interaction of distance and population density captures the heterogeneity in consumers' preferences with respect to distance to shops. ε_{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 people in the population. u_{i0} allows the utility of consumer i shopping from the outside option to depend on its location characteristics. For example, more populated areas have more outside options and, thus, higher utility of not shopping at any Blue or Red stores in the choice set. This attempts to control for other competitors that Blue and Red firms face in the market.

Block group i's choice set is defined as the Blue and Red stores within r_i miles of i's location. r_i is a function of population density, $25 \times (1 + (median(popden) - popden_i)/median(popden))$. r_i takes on values 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, in terms of the farthest Blue or Red store to which they are willing to travel, across areas with different population density. r_i increases as population density decreases. In other words, people living in rural areas might be more willing to travel farther 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 can be written as

$$R_{jl} = \sum_{i:d_{il} \leq r_i} \lambda \cdot p_{ijl} \cdot n_i, \tag{3.1}$$

where λ is average spending per consumer and n_i is the total population of block group i. Ideally, λ can depend on consumer characteristics w_i. But the data are not detailed enough to identify $\lambda(w_i)$. This is because sales are observed only at the store level, and 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 λ is the average spending per consumer across the nation. In one of the empirical specifications, λ is allowed to depend on whether store j sells general merchandise only or both general merchandise and groceries.[11] Results do not change very much. (See Section 4.1 for details.)

3.3 Firms' entry decision

In this section, I introduce the dynamic discrete choice model used to study firms' entry decision. I describe two-stage budgeting, as well as the clustering algorithm, and show how they can be applied to make the model tractable. I give an overview of the model in subsection 3.3.1; present the details of the model in subsection 3.3.2; discuss two-stage budgeting in subsection 3.3.3; derive the sufficient conditions for two-stage budgeting to be valid namely, separability–in subsection 3.3.4; and explain the clustering algorithm in subsection 3.3.5.

3.3.1 Overview of multi-store chain's entry model

There are three features in the multi-store chain's entry model. The first is that firms maximize payoffs over all stores instead of independently for each individual store. This is important because the stores are under shared corporate ownership instead of franchise agreements. Furthermore, the spatial interdependence across stores, such as cannibalization and cost sharing, would not be accounted for if stores were treated independently. The second feature is that firms are forward-looking. Given that demographics and distribution networks change over time, it is reasonable to assume that firms maximize the sum of expected current and future payoffs. This is also necessary when examining preemptive incentives. Previous studies, such as Holmes (2011), show that dynamic consideration is important for the entry decisions of retailers such as Walmart. The third feature is that there are strategic interactions between firms, which is supported by the fact that Blue and Red firm compete in many markets, as shown in Figures 1 and 2. This is also necessary for studying firms' preemptive entry behavior. Findings in Jia (2008) provide more evidence of strategic interactions between discount retailers. To incorporate these three different features, I study firms' entry decisions using a dynamic discrete choice game model in which decisions are made at the firm level.

[11] As described in Section 2.2, the sales data of stores that sell both general merchandise and groceries include the sales of general merchandise only.

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 store location decision, each firm takes into account the following factors in both current and future periods: consumer demand, distribution networks, local wage and rent, its own potential store openings, and its opponent's possible store openings. The decision is made at the firm level instead of individual store level. Budget constraints determine the number of new stores to be opened by each firm. Firms move sequentially in each period, with Blue firm moving first.

The large number of store openings and possible store locations of the two firms lead to the very large state space in the dynamic game. As a result, the payoff optimization problem has high computational complexity for the chain store managers, as well as for the econometrician. Therefore, it is necessary to reduce the size of the state space, and the desirable tools are those that mimic the way chain store managers solve the problem in reality.

3.3.2 Two-player discrete choice game

Let π^l_{jt} be the static profit of firm j's store l in period t.

$$\pi^l_{jt} = \mu_j R^l_{jt}(s_t) - w^l_t E(R^l_{jt}(s_t)) - r^l_t L(R^l_{jt}(s_t)) - \psi_j D^l_{jt} - \alpha_j x^l_t, \tag{3.2}$$

where μ_j is the gross margin of firm j. $s_t = (s_{jt}, s_{-jt})$, where $s_{jt} \in \{0,1\}^L$, indicating that 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. Following Holmes (2011), labor cost and land cost are modeled as variable costs. 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 land size of the store. $\psi_j D^l_{jt}$ is the distribution cost, where ψ_j is per unit distribution cost and D^l_{jt} is the distance to the closest distribution center. $\alpha_j x^l_t$ is the fixed cost, which depends on population density around store l, x^l_t. Each period t is a quarter. The firm-level static profit is

$$\pi_{jt} = \sum_{l=1}^{L} s^l_{jt} \left\{ \mu_j R^l_{jt}(s_t) - w^l_t E(R^l_{jt}(s_t)) - r^l_t L(R^l_{jt}(s_t)) - \psi_j D^l_{jt} - \alpha_j x^l_t \right\}, \tag{3.3}$$

where the sum is over all locations $\{1, ..., L\}$ and s_{jt}^l is the l^{th} component of s_{jt}.

Next, I introduce the value function of firm j. For simplicity, I assume that moves are sequential and that Blue firm moves first.[12] 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 available locations in period t–in other words, all possible locations minus those taken by the two firms before period t–i.e., $L_t = L/\{l \in L : s_{jt}^l + s_{-jt}^l = 1\}$. Let a_{jt} be the vector $\{a_{jt}^l\}_{l=1}^L$; then, $s_{jt+1} = s_{jt} + a_{jt}$. 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 $s_{jt} = (s_{jt}, z_{jt}, B_{jt})$. $s_t = (s_{jt}, s_{-jt})$ is the state variable in period t. Firm j's value function in period t is

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

s.t.

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

where $\mathbb{A}_t = \{0, 1\}^{L_t}$ is the choice set in period t. The expectation is over a cost shock η_{jt}^l of opening a store at location l. η_{jt}^l are i.i.d. across locations and time periods. $P(s_{-jt+1} | s_{jt+1}, s_{-jt})$ is the transition probability of j's opponent in period t. $f(a_{jt}^l)$ is a budget function that I will discuss in more detail later in this section. β is the discount factor.

Each period, firm j chooses the optimal entry decision $a_{jt} \in \mathbb{A}_t$ to maximize the sum of expected profits $\mathbb{E}\pi(s_{jt} + a_{jt}, s_{-jt})$ and the continuation value $\beta \sum_{s_{-jt+1}} \mathbb{E}V(s_{jt} + a_{jt}, s_{-jt+1}) P(s_{-jt+1} | s_{jt+1}, s_{-jt})$. The distribution of η_{jt}^l is common knowledge, but its realization is private information. Firm j's strategy σ_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

[12]Ideally, I would like to solve the game with Red firm moving first, but it is computationally demanding.

(3.5). $f(a^l_{jt})$ indicates both the financial and the management cost of firm j opening a store at location l in period t. It includes the actual costs of acquiring land or building infrastructure of a store, 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 that firms behave. According to the interviews conducted with chain store managers, 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 were expanding in the sample period between 1985 and 2001, and the financial constraints can often be a serious consideration when firms are expanding. Figure 5 plots the book value of total assets of Blue firm and Red firm in the sample period, in their respective colors. The figure shows that Blue firm's book value of total assets increased quickly in this period. Thus, it is likely that the financial constraint that Blue firm was facing was substantial at the beginning of this period but was 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 because of its financial costs and management costs. For simplicity, I assume hereafter that $\sum_{l=1}^{L_t} f(a^l_{jt}) = \sum_{l=1}^{L_t} a^l_{jt} \leq B_{jt}$–i.e., the total number of new stores that 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 that Blue firm 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. First, for the purpose of studying preemptive incentives, it is reasonable to focus on locations that 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 in Section 4.2, may lead to biased results. The drawback of defining the set of potential locations as the observed stores by the end of the sample period is that firms' decisions are very much affected by the limited choice set towards the end of the sample period. To avoid this problem, I leave out the last two years of data and include only observations between 1985 and 1999 as the study sample.[13] The choice to omit the two years will be explained in Section 4.3.

Modeling entry decisions at the 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 model intractable. Each period, firms choose 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 was 1262. The size of the state space bounded below by $\binom{1262}{27} \approx 10^{35}$.

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 of the model described above. Two-stage budgeting refers to a type of model used in studying income allocation problems. In these models, consumers first allocate a given amount of total expenditures to categories of goods and then optimize consumption within each category, conditional on the amount designated to that category of goods (Gorman, 1971).[14] I apply the same idea to chain stores' entry problem. Store locations are similar to goods in the consumption problem. Let $\{1, ..., P_{M_t}\}$ 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 $P_{mt} \in \{1, ..., P_{M_t}\}$,

[13] Locations of the stores opened in 2000 and 2001 are included in the choice set L.

[14] Note that only the separability conditions in Gorman (1971) are needed for two-stage budgeting to be valid; the conditions for constructing a price index are not necessary.

$$V(s_{jmt}, s_{-jmt}|B_{mt}) = \max_{a_{jmt}\in\mathbb{A}_{mt}} \Big\{ \mathbb{E}\pi(s_{jmt}+a_{jmt}, s_{-jmt}) + \beta \sum_{s_{-jmt+1}} \mathbb{E}V(s_{jmt}+a_{jmt}, s_{-jmt+1}|B_{mt}) \cdot P(s_{-jmt+1}|s_{jmt+1}, s_{-jmt}) \Big\} \quad (3.6)$$

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_{mt} of the partition $\{1, ..., P_{M_t}\}$. Equation (3.6) corresponds to solving the consumption problem within each category of goods given the amount of income allocated to that category. In the next stage, firm j solves for the optimal budget $\{B_{j_1 t}, ..., B_{j_{M_t} t}\}$ for each element P_{mt} of the partition:

$$\sum_{m=1}^{M_t} \mathbb{E}V(s_{jmt}, s_{-jmt}|B_{jmt}) \quad (3.7)$$

s.t.

$$\sum_{m=1}^{M_t} B_{jmt} \leq B_{jt}.$$

This corresponds to solving for the optimal income allocation for each category of goods in the consumption problem.

With two-stage budgeting, firms solve two smaller optimization problems, Equation (3.6) and (3.7) instead of (3.4). This decentralization greatly reduces the size of the state space in estimation. When estimating the parameters in Equation (3.4), one can condition on the the observed budget for each element of the partition $\{B_{jmt}\}_{m=1}^{M_t}$, and only solve Equation (3.6). The size of the state space is then reduced to $\sum_{m=1}^{M_t} \binom{|P_{mt}|}{\sum_{l\in m} a_{jt}^{l}}$.

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

However, it is not clear whether solving Equation (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 in (3.4). Then, I derive sufficient conditions on the primitives of the model such that the separability conditions are satisfied. I define a <u>market</u> to be an element of the partition, P_{mt} and I show that the solution to the two-stage budgeting problem is optimal if separability across markets is satisfied.

3.3.4 Separability conditions in a two-player Markov game

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

$$max_{\sigma_{j(s)}} \pi(s_j + a_j, s_{-j} + a_{-j}) \text{ s.t. } \textstyle\sum_{l=1}^{L} a_j^l \leq B_j. \tag{3.8}$$

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

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

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

Definition 1 *Locations* $\{1, \cdots, L\}$ *are* <u>*separable*</u> *in the partition* $\{P_1, \cdots, P_M\}$ *if*

$$\frac{\Delta\mathbb{E}\pi(s_j, s_{-j}, l)}{\Delta\mathbb{E}\pi(s_j, s_{-j}, h)} \perp (s_j^k, s_{-j}^k), \ \forall l, h \in P_{m_l}, \ \forall k \in P_{m_k}, l \neq k.$$

In other words, if the ratio of the expected marginal profits of opening stores in any two locations in a market does not depend on the state variables in another market, locations are separable with respect to markets. This 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 σ_j. Note that $\sigma_j(s)$ 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 σ_j^* be the best response of j given opponent's strategy σ_{-j}, and a_j^* be the corresponding optimal action at state (s_j, s_{-j}).

Definition 2 *Firm j's strategy σ_j^* is separable in the partition $\{P_1, \cdots, P_M\}$ if for given σ_{-j}, $\exists \sigma_{j1}^*, \sigma_{j2}^*, \cdots \sigma_{jM}^*$ s.t.*

$$\sigma_{jm}^*(s_{jm}, s_{-jm}, B_m) = \sigma_j^{*m}(s_j, s_{-j}, B),$$

where $\sigma_j^* = \left(\sigma_j^{*1}, \cdots, \sigma_j^{*M}\right)$, $B_m = (B_{jm}^*, B_{-jm})$, $B_{jm}^* = \sum_{l \in P_m} a_j^{*l}$, $B_{-jm} = \sum_{l \in P_m} a_{-j}^l$, $\forall m = 1, \cdots, M$, and $\sum_{m=1}^M B_m = B$.

In other words, σ_j^* is separable if each component σ_j^{*m} can be written as a function σ_{jm}^* which, depends only on the state variable in the partition (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 σ_{-j}^* is separable, then j's optimal strategy σ_j^* is separable.*

See Appendix I for the details of the proof. Theorem 1 states that if locations $\{1, \cdots, L\}$ are separable, and one firm is playing a separable strategy, then it must be optimal for the other firm to play a separable strategy. 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 $R_j^l(s_t)$; thus, $R_j^l(s)$ 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 η_j^l.

Theorem 2 *The locations $\{1, \cdots, L\}$ are separable in partition $\{P_1, \cdots, P_M\}$ if the expected profit function $\mathbb{E}\pi(\cdot)$ satisfies the following conditions:*

1. *$R_j^l(s)$ 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_j^k and x_j^k, where $k \in P_n$,$m \neq n$;*

3. *η_j^l are independently distributed across markets.*

See Appendix I for the proof. By Equation (3.1), it is clear that if $\nexists i$, s.t.

$$p_{ijl} > 0, p_{ijk} > 0, l \in P_m, k \in P_n, m \neq n, \tag{3.9}$$

then the first condition in Theorem 2 holds. In other words, if there does not exist consumer i that shops at 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, however, if the cost structure is different. For example, if the distribution center has a capacity constraint, and per unit cost of distributing depends on the number of stores the distribution center serves, separability does not hold across locations that share the same distribution center but belong to different markets. The third condition is satisfied by the i.i.d. assumption on the cost shock η_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 those in 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); and 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 and decentralize firms' entry decisions, a sufficient condition is that markets are separable, as discussed in the previous section. In this section, I explain how the U.S. national market can be divided into smaller separable markets. To divide the market, I apply a clustering algorithm designed for the separability conditions to hold. I first introduce the objective function for the clustering algorithm, and then explain the steps of the algorithm to find an optimal partition of store locations given this objective function. Each element of the partition is defined as a market. Results are postponed to Section 4.2.

Given the cost structure, the main condition that needs to hold for markets to be separable is that store 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 far enough away from each other that no consumer has both stores in its choice set. Clearly, the difficulty in dividing the market arises when two markets are next to each other, and consumers living close to the border of the markets are willing to shop at stores in both markets. In reality, two neighboring stores rarely share zero customers, except in areas where population density is extremely low and stores are very far from each other. I define the objective (or loss) function of the clustering algorithm to measure how far away a given partition is from being truly separable. The algorithm is applied to find the solution that minimizes this distance. Ideally, the distance is zero and all markets are separable. In reality, the clustering algorithm finds the market definition that is the closest to a separable partition of store locations.

Define the objective (or loss) function as the following:

$$min_{\{P_1,\cdots,P_M\}} \sum_{l=1}^{L_t} [R^l(s,\omega) - R^l(s_m,\omega_m|l \in P_m)]^2, \tag{3.10}$$

where L_t[15] is the set of potential locations in period t, and $R^l(s,\omega)$ is the revenue of store l, which depends on the state variable s and the determinants of demand ω, which include demographic characteristics and store characteristics. Note that both s and ω are vectors that contain information about the entire U.S. national market. The second term, $R^l(s_m,\omega_m|l \in P_m)$, is also store l's revenue, but it is computed using information on existing store locations and demand data only in the partition P_m. In other words, it is the revenue of store l when 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. That is, if l and h share any customers, those customers are restricted to shopping at only one of the two stores. Consumers in block

[15] The t subscript is kept 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.

groups that have both l and h in their choice set are assigned to the market to which their closest store belongs,.[16] If P_m is truly separable from the rest of the markets, then $R^l(s, \omega) - R^l(s_m, \omega_m | l \in P_m)$ is zero for all $l \in P_m$. Therefore, the sum of squared differences between $R^l(s, \omega)$ and $R^l(s_m, \omega_m | l \in P_m)$ indicates how far off a partition is from each of its elements being truly separable, or the loss of spatial interdependence by assuming that markets are separable. The solution to Equation (3.10) is the optimal partition that minimizes this loss.

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 approximated. Start with $M = 2$. Apply a greedy algorithm that 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, the greedy algorithm is more suitable than other algorithms, such as the spectrum algorithm which has the advantage of speed but assumes additional structures of the problem. I describe the greedy algorithm and the stopping criterion in the remainder of this section.

The greedy algorithm finds the (approximated) 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 represents a store location, and the edge between a pair of dots means that the two stores share customers. The task is to cut off a set of 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. Two features of the problem are important for the setup 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 the markets. Second, the edges between stores are weighted. The weight is the amount of interdependence between two stores and is decided by the consumer demand and the objective function. Since the weight varies across edges, one cannot simply minimize the number of broken edges in a graph to find the optimal partition.

The algorithm follows four steps.

[16] Another way to group consumers is to assign them to the market to which their most preferred store belongs. Results are similar.

1. For a given partition $\{P_1, \cdots, P_M\}^t$, find all locations 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 that 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 locations fixed. Call the new partition $\{P_1, \cdots, P_M\}^{(t+1)}$.

3. Repeat steps 1 and 2 until the algorithm converges.

4. Repeat steps 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 a given store location(s) belong to one partition, the store(s) itself cannot be in a different partition. In other words, if $\forall h \in C_l$, $h \in P_m$, 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 the incremental change of loss from M to $M+1$ partitions is greater than or equal to one percent of revenue $R^l(s, \omega)$ for any store location l.[17]

4 Estimation and Clustering

4.1 Demand estimation

The demand model is estimated using a maximum likelihood framework. Following Holmes (2011), the discrepancy between the model and the data is assumed to be a measurement error that follows a normal distribution. Denote the measurement error by ϵ_{jl} and the observed sales of store l by R_{jl}^{obs}. Then,

$$ln(R_{jl}^{obs}) = ln(R_{jl}) + \epsilon_{jl},$$

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,

[17] It also happens to be the point at which the incremental change of loss increases dramatically in many cases.

which is \$2444 per year. This is close to the estimate in Holmes (2011), which is about \$2150 per year in 2007 dollars. The coefficient on population density is positive and significant. The coefficient on distance is negative and significant. Column 2 presents a different specification with a dummy variable indicating whether a store sells both general merchandise and groceries or general merchandise only. The results are similar.

The comparative statics presented in Table 7 illustrate the effects of distance to shops 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 the Blue store stays at two miles. From left to right, choice probabilities decrease as population density increases, showing that the utility of choosing the outside option increases as population density increases.

4.2 Clustering results

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

Since the set of potential locations L_t changes in each period, the partitioning of markets is conducted every period. Note that the existing stores are not partitioned because they are no longer in the choice set of the firms. Consequently, the spatial interdependence among existing stores and between existing and potential stores is fully accounted for. As a result, in the empirical analysis, market definitions are different in each period. Although this assumption has the drawback of not allowing any market-level unobservables that can be controlled for using market fixed effects, it has the advantage of being close to the way that managers make decisions in reality. According to interviews with managers and consultants in the industry, when firms evaluate a potential store location, they first define a trade area around it. The trade area is where demand is likely to come from and where the main competing stores, including the firms' own stores and competitors', are located. Naturally, the trade area varies across time as new stores are opened each period. As a result, the clustering procedure can also be viewed as estimating the trade areas, which is important to firms' decision making; however, the econometrician has no knowledge of it.

Figure 7 is a map of the Northeast, Mid-Atlantic and (part of the) Southeast U.S.–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. For example, the five green squares in the upper right corner are in one market, while the two blue squares to their left are in a different market. The yellow patches on the map are the CBSAs. This map compares market divisions defined by the clustering algorithm to the CBSA units. In a few cases, using the CBSA to define a market is not very different from the clustering results. For example, the single black dot in the upper right corner is the only store in its market after clustering the store locations. In this particular case, defining the market as the CBSA in which the store is located is not a bad idea since there are no other stores nearby, and most consumers shopping at this store are likely to live in this CBSA. However, in most cases, defining a CBSA as a market can be misleading. For example, the five blue dots to the lower left of the black dot are defined as being in one market by the clustering algorithm, while they are located in three different CBSAs. Dividing the stores into three different markets would also be misleading since they are very close to each other, and at least two of them are right on the border of two different CBSAs. Consumers do not confine themselves to shopping within CBSA boundaries, so it is not reasonable to define markets by the CBSAs in which the stores are located. On the other hand, defining markets by applying the clustering algorithm minimizes the restriction of market definition on consumers' shopping behavior.

Table 8 presents measures of goodness of the clustering results in comparison to the CBSA market definition. The measures are computed using all L_t locations in the third quarter of 1997. There are 241 locations. First, the total loss by clustering as a fraction of total revenue–i.e., $[\sum_l |S(l) - S(l \in r^*)|]/[\sum_l |S(l)]$–is less than 0.001 percent. The maximum store-level loss as a fraction of revenue is also reasonably small, 0.5 percent. 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 percent. On the other hand, using CBSAs as markets would lead to undesirable results. For example, the maximum store-level loss is 53.0 percent, about a hundred times higher than that of clustering. Therefore, clustering is the more desirable way to divide markets when interdependence across stores is an important concern.

4.3 Cost estimation

In this section, I explain the estimation of the cost function. Due to the non-stationary nature of the game, which is explained below, the estimation takes a different approach from the two-stage procedure proposed by BBL or POB. The approach includes two parts. First, it computes the value function at each possible state using a 'rolling window' approximation. Second, the game is solved using backwards induction and the computed continuation values. The advantage of this approach is that it is more suitable for the counterfactual analysis of interest–which is to quantify preemptive incentives–and, therefore, requires evaluating the value functions at those states that are not observed in the data. Moreover, computing the value functions using the 'rolling window' approximation mimics how managers make decisions in reality.

First, I describe the estimation of the parameters that does not require solving the dynamic game. Recall that the 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). μ_j is computed using information in firms' annual reports. The average gross margin for the Blue and Red firms is 0.24 and 0.22, respectively. The amount of labor and land of a store is assumed to be proportional to revenue:

$$E(R^l_{jt}) = \mu_E R^l_{jt},$$

$$L(R^l_{jt}) = \mu_L R^l_{jt}.$$

μ_E is calibrated using labor costs listed in firms' annual reports. μ_L is computed using census data and county property tax data of a subset of stores. See Appendix A in Holmes (2011) for details.

Next, I explain the dynamic game estimation. The parameters of interest is per unit distribution cost ψ_j and the fixed cost α_j of the Blue and Red firms. Recall that 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 also solve the upper-level optimization problem to get the optimal number of new stores allocated to each market.

Most dynamic game estimation methods in the literature follow a two-stage procedure such as those in BBL or POB. However, the current problem does not fit in the two-stage estimation framework for the following two reasons.

First, in the two-stage estimation 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 setting, however, no state is observed more than once for the following reasons. First, both firms are expanding in the sample period, and the total number of stores increases as new stores are added every period. Dividing the national market into smaller markets solves the problem by creating repeated observations across markets. However, the set of potential and existing locations across markets is not treated as interchangeable, which is the second reason that the non-parametric estimation of choice probabilities is not possible. The nature of the dynamic problem in the current setting is non-stationary for the above two reasons. To convert the problem into one that fits the two-stage estimation framework, one solution would be to categorize the observed states into groups by some similarity measure and treat each group as a single state. This is not desirable because the state variable contains rich geographic information that is important for studying preemptive incentives. Location characteristics contained in the state variable include the number of nearby stores belonging to the same firm and to the competitor, the distance to those stores, the distance to the distribution centers, and consumer demographic information. Those variables can be crucial for identifying preemptive incentives. Pooling them into groups may lead to loss of information and, therefore, bias the results.

Second, even if the estimation can be done using a 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 preemptive incentives to entry are partially removed, it is 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. In other words, the value function needs to be evaluated at a different set of states in the counterfactual than those observed in the sample. 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 set of parameter values, 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 a search in the parameter space to find the set of values that maximizes the likelihood of the observed choices. 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 difficult to compute. Next, I describe how the value function can be approximated using a 'rolling window' method.

In each period t and each market m, firm j chooses B_{jmt} locations to open new stores from L_{mt} potential locations. The set of potential locations L_{mt} includes all the locations that the Blue and Red firms 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 that the terminal value equals $\mathbb{E}V(s_{mT})/(1-\beta)$. β 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 eight periods after T instead of completely static. The implicit assumption is that firms do not foresee any more entry or change in demographics after $T+8$. This is a limitation, but no more data are available.

Second, to compute the continuation value for each of the $\binom{L_{mt}}{B_{jmt}}$ states, the value function needs to be evaluated at each of the possible states in the future between t and $T+8$. This is computationally infeasible. For example, at period $t=1$, $L_t = L = 3123$. Assuming that clustering can reduce the market size to 30,[18] the number of value functions that need to be computed is

$$\binom{L_m}{B_{1m1}} \cdot \binom{L_m - B_{1m1}}{B_{2m1}} \cdots \binom{L_m - \sum_{\tau=1}^{T} \sum_j B_{jm\tau}}{B_{1mT}} \cdot \binom{L_m - \sum_{\tau=t}^{T} \sum_j B_{jm\tau} - B_{1mT}}{B_{2mT}}$$
$$= \binom{L_m}{B_{1m}} \cdot \left(\binom{B_{1m}}{B_{1mT}} \cdot \binom{B_{2m}}{B_{2mT}} \cdot \binom{B_{1m} - B_{1mT}}{B_{1m(T-1)}} \cdot \binom{B_{2m} - B_{2mT}}{B_{2m(T-1)}} \cdots \right), \quad (4.1)$$

where B_{jm} is the total number of new stores that belong to firm j in market m, $j = 1, 2$. Assume that 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 method to compute the value function is necessary for both the manager of the firm and the econometrician.

As discussed above, changing the state variables to reduce the size of the state space is not a desirable approach. One can also reduce the size of 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 $T+8$, 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:

$$V(s_{jmt}, s_{-jmt}) = \max_{a_{jmt} \in \mathbb{A}_{mt}} \left\{ \sum_{\tau=t}^{t+8} \frac{\beta^{\tau-t}}{(1-\beta)^{\mathbb{1}\{\tau=t+8\}}} \sum_{s_{-jm\tau}} \sum_{s_{jm\tau}} \mathbb{E}V(s_{jm\tau}, s_{-jm\tau}) \right.$$
$$\left. P(s_{jm\tau} | s_{jmt} + a_{jmt}, s_{-jmt}) P(s_{-jm\tau} | s_{jmt} + a_{jmt}, s_{-jmt}) \right\} \quad (4.2)$$

s.t.

$$\sum_{l=1}^{\bar{L}_{mt}} a_{jmt}^l \leq B_{jmt},$$

[18] This is already unrealistic. It implies that the US national market is divided into 100 markets; therefore, the loss from clustering must be substantial.

where $\bar{L}_t$ is the set of locations entered by the Blue and Red firms between t and $t+8$, and $(1-\beta)^{\mathbb{1}\{\tau=t+8\}}$ is a scaling factor for terminal period $t+8$. This reduces the computational burden dramatically. Clustering is now done over the set of locations $\bar{L}_t$, which includes all potential locations between t and $t+8$. The maximum size of the potential market, $\bar{L}_{mt}$, is 12[19] after taking out a few city centers.

Moreover, the approximation is close to the way that managers make decisions in reality. According to the interviews conducted with managers and consultants, managers usually make two types of plans, a two-year and a five-year plan. While the five-year plan sets general long-term goals, managers typically have a fairly good idea of how many stores they are planning to open in the next two years, which corresponds to eight periods in the model, and where the potential locations are. 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.

On the other hand, the approximation imposes two restrictions on the firms' optimization problem. First, since the optimization stops at $t+8$, any change in the state variable after $t+8$–and, thus, any change in the continuation value–is not taken into account. The number of possible paths that need to be evaluated becomes

$$\binom{L_{mt}}{B_{1mt}} \cdot \binom{L_{mt}-B_{1mt}}{B_{2mt}} \cdots \binom{L_{mt}-\sum_{\tau=t}^{t+7}\sum_j B_{jm\tau}}{B_{1mt+8}} \cdot \binom{L_{mt}-\sum_{\tau=t}^{t+7}\sum_j B_{jm\tau}-B_{1mt+8}}{B_{2mt+8}}. \tag{4.3}$$

Second, the possible paths that firm j is optimizing over between t and $t+8$ are further restricted due to the reduced choice sets by the rolling window. All the L_t terms in (4.3) become $\bar{L}_t$, 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 the observed locations $\bar{L}_t$ only–that is, any location entered by either firm after period $t+8$ is not considered a potential location in period t.

Finally, I conduct a grid search through the parameter space and use maximum likelihood to estimate the parameters $\{\psi_j, \alpha_j\}$, $j=1,2$. Assuming that the cost shock of each action follows a type I extreme value distribution, the choice probabilities $P(s_{jmt+1}|s_{jmt}, s_{-jmt})$ have a closed-form solution. For a given parameter value, the game in (4.2) is solved for

[19] Note that this does not include the existing stores.

each market, period, and firm. The choice probabilities are matched to the observed choice by:

$$max_{\psi_j,\alpha_j} \sum_m \sum_t \sum_j log\left(Pr(a_{jmt}|s_{jmt}, s_{-jmt}, \psi_j, \alpha_j)^{Y(a_{jmt})}\right), \tag{4.4}$$

where

$$Pr(a_{jmt}|s_{jmt}, s_{-jmt}, \psi_j, \alpha_j) = \frac{exp(\mathbb{E}V(s_{jmt} + a_{jmt}, s_{-jmt}))}{\sum_{a_{jmt}\in\mathbb{A}_{jmt}} exp(\mathbb{E}V(s_{jmt} + a_{jmt}, s_{-jmt}))},$$

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

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

4.4 Estimation results and interpretation

Table 9 presents the estimation results of distribution cost and fixed cost by firm. The distribution cost per thousand miles is 1.61 million dollars for Blue firm and 0.68 million dollars for Red firm. Using industry sources, Holmes (2011) estimates Blue firm's trucking cost of distribution to be 0.8 million dollars per thousand miles. Using a single agent dynamic model, he estimated the total distribution cost to be around 3.5 million dollars per thousand miles. My estimate of 1.61 is in the interval of $[0.8, 3.5]$ and closer to the industry source of trucking costs than Holmes' estimate. Holmes interprets his estimate of distribution costs 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 effect in my results is due mainly to the fact that the analysis takes into account the strategic interactions between firms. 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 if, as will be shown below, if the competitor (Red firm) has an urban advantage. Therefore, when taking into account the strategic interactions between the two firms, the overall economies of scale become smaller.

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

The average fixed cost of operating increased by 0.43 million dollars and 0.15 million dollars per year for Blue and Red firm, respectively, when population density increased from 250 (25th percentile) to 700 (50th percentile) thousands of people per circular area with a five-mile radius in 1990. The average fixed cost per store in 1990 was about 1.84 and 0.62 million dollars, or four percent and five percent of sales, for Blue firm and Red firm, respectively. Since 0.62/1.84 is less than 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 by bootstrapping over markets. The calculation does not include estimation errors of demand in the first stage or clustering errors in the second stage.[20]

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 partially remove the preemptive motives using a one-period deviation method and compare firms' optimal response to that of the equilibrium outcome with preemption. In other words, the analysis is designed to answer the following question: what would the optimal entry decisions 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 producer surplus of about one million dollars per

[20] I propose a simulation method to compute standard errors, taking into account the errors in the first and second stages of the estimation. See Appendix II for details.

store, measured by the combined sum of the two firms' current and future profits.

It is a difficult task 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 the model. 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 future state of j and the source of dynamic optimization for a single agent; and $\{s_{-j\tau}\}_{\tau>t}$, which is the opponent's possible state in the future and the source of preemptive incentives. Moreover, as preemption is a motive rather than an action, the econometrician cannot directly observe it. Given the complex setting in which preemption arises and its unobserved nature, it is often difficult 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 is close to the original one, such that payoffs are comparable. Thus, solving for a different equilibrium in which firms do not have incentives to preempt 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 other dynamic considerations. Moreover, the analysis stays in the current strategy space, which allows me to compute the loss of producer surplus due to preemption. I define the preemptive incentives of firm j entering a location l to be the change in $Pr(a^l_{jt}|s_t)$ in response to $Pr(a^l_{-jt'}|s_t)$ in equilibrium, where $Pr(a^l_{jt}|s_t)$ is the choice probability of firm j entering location l at state s_t in equilibrium, and $Pr(a^l_{-jt'}|s_t)$ is the same probability for $-j$ in period t', $t' > t$. That is, 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 $Pr(a^l_{jt}|s_t)$ when $Pr(a^l_{-jt'}|s_t)$ is set to zero. The idea is the following: for each of Blue firm's observed choices, remove those choices from Red firm's choice set for one period. Thus, Blue firm knows that Red firm would not be allowed to enter those locations for one period. Then, I investigate whether Blue firm has profitable deviations by delaying entry at those locations. Specifically, I solve the following equation to compute the choice probabilities for Blue firm without preemption,

$$V(s_{jmt}, s_{-jmt}) = \max_{a_{jmt}\in\mathbb{A}_{mt}} \left\{ \sum_{\tau=t}^{t+8} \beta^{\tau-t} \sum_{s_{-jm\tau}} \sum_{s_{jm\tau}} \mathbb{E}V(s_{jm\tau}, s_{-jm\tau}) P(s_{jm\tau}|s_{jmt}+a_{jmt}, s_{-jmt}) \right.$$
$$\left. P(s_{-jm\tau}|s_{jmt}+a_{jmt}, s_{-jmt})|a_{-jmt} \neq a^{obs}_{jmt} \right\} \quad (5.1)$$

s.t.

$$\sum_{l=1}^{\bar{L}_{mt}} a^{l}_{jmt} \leq B_{jmt},$$

where a^{obs}_{jmt} is Blue firm's observed choice in period t. Restricting $a_{-jmt} \neq a^{obs}_{jmt}$ 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. Note that the market-level budget constraint is held fixed, so Blue firm is not fully optimizing. Relaxing $\bar{L}_{mt}$ would only lead to a 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 size of preemptive incentives measured by the above procedure is a lower bound of the true value.

For 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, Red firm would switch away to other markets, inducing a weaker presence in the current market. This would lead to a 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 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 firm's choice set for one period. For those 425 locations where preemption is observed, the 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.

Second, I study the loss of producer surplus due to preemption. Since I do not have enough data to conduct total welfare analysis, I compute efficiency loss from firms' perspectives by examining the change in the sum of expected current and future profits of the two firms. Results are shown in the last two rows of Table 10. Out of the 425 locations, at 392, the expected total payoff of Blue firm 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 around 397 million dollars. On average, the loss of producer surplus per location is 1.01 million dollars, which is about the same as the annual profit of a small to median-sized store. For the reasons discussed above, it is a lower bound of the producer surplus loss as a consequence of preemption. The market-level budgets are held fixed and firms are not fully optimizing. From the producers' perspectives, preemption results in a substantial loss of efficiency.

Next, I investigate why evidence of preemption is found at 425 locations but not at others. One reason is that preemption may not always be profitable. I compare the characteristics of the two sets of locations in Table 11 and refer to them as preemption and non-preemption locations. The differences are not all significant but are qualitatively consistent with the intuition and the reduced-form evidence shown in Section 2.3. First, the distance to Blue firm's distribution center is shorter for the non-preemption stores than for the preemption stores. The opposite is true for distance to Red firm's distribution centers. 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 those locations would mean giving up current profits; hence, preemption is more likely to be observed at those locations where current profits are low, which are those that are a longer distance from Blue's distribution center. Therefore, in the current one-period deviation experiment, preemption is more likely to be observed at locations that are closer to Red's distribution center and farther away from Blue's distribution center. Moreover, preemption stores also tend to locate in areas with higher store and population density. This is also consistent with the fact that those are the areas where Red store is more likely to enter.

Finally, I compute the one-period deviation for Red firm. In this experiment, Blue firm is forced to stay away from its observed choices for one period, and Red firm is given the opportunity to enter those locations for one period. I examine whether Red firm would choose to enter those locations and, if so, how much its payoff would increase. Note that the analysis is computationally feasible 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. Since Red firm could have re-optimized across markets, the effect of preemption measured is a lower bound, as in the previous experiment.

The results are presented in Table 12. At about one third of the locations, it would be profitable for Red firm to enter right away if Blue firm stayed out of those locations for one period. The average probability of entry is 0.77. The average payoff increase is 2.99 million dollars for each location, which is about a large store's one-year profit. The analysis of the two counterfactual scenarios indicates that the impact of preemptive incentives on entry decisions is significant from the perspectives of both firms. [21]

In the current literature, there are two methods for identifying preemption using empirical analysis. The first method, used in Schmidt-Dengler (2006), quantifies 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 resulting equilibrium is compared to the original equilibrium, in which firms are allowed to re-optimize every period without pre-commitment, and the difference is interpreted as preemption. The reason this approach is not satisfying is twofold. First, it does not exclude preemption completely. The pre-commitment game 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 strategy space and lead to completely different equilibria, it is difficult to compare payoffs with the original equilibrium.

[21] The counterfactual analysis can be generalized to multiple locations and multiple periods, which would capture a larger effect of preemption.

The second method is to solve a single agent's dynamic problem, as in Igami and Yang (2014). Applying this method in the current setting, Blue firm solves for the optimal strategy, assuming that Red firm does not respond. From Blue firm's perspective, Red firm is part of nature and, therefore, cannot be preempted. This is equivalent to Blue firm solving for a single agent's dynamic optimization problem, while Red firm plays according to the original equilibrium actions. The difference between the solution and the original equilibrium is interpreted as preemption, according to Igami and Yang (2014). 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 between the two firms. Moreover, the assumption that Red firm sticks to the original equilibrium play regardless of Blue firm's deviation is not well motivated.

6 Counterfactual II: Subsidy Policy after Red Firm Exits

Red firm started exiting in many markets by closing stores in the early 2000s. 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), and local governments have proposed subsidies to induce 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.[22] Rolling Meadows, IL managed to subsidize Blue firm to enter after the local Red store closed.[23] However, large amounts of retail space remained empty for years after Red firm exited. Rockledge, FL and Indiana Harbor Beach, FL are two examples.[24] The former retail space of the Red store in Rockledge stayed empty for 12 years. Thus, local governments' subsidy policies have not been effective. Therefore, it is important for policy makers to better understand whether and how government subsidies affect retailers' entry decisions. The current entry model can be used to answer this question. In order to evaluate the subsidy policies, it is also important to consider consumer welfare loss due to store closings. Although I do not have enough data to conduct total welfare analysis, I can compute the increase of

[22]Source: http://www.huffingtonpost.com/2012/01/26/sears-closes-cities_n_1231326.html
[23]Source: goodjobfirst.org.
[24]Source: http://www.floridatoday.com/story/money/business/2014/07/27/kmart-goes-next/13197001/

consumer drive time to a discount retail store due to Red store closings and compare it to the size of the observed subsidies. This section attempts to answer those two questions by computing the size of subsidy needed for Blue firm to enter an ex-Red location and deriving the welfare loss of store closings to consumers due to higher shopping costs.

First, I compute the expected payoff of Blue firm entering each ex-Red location and compare it to the expected payoff from 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 size of 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 to Blue firm. Then I compute the size of subsidies separately for two sub-periods: 2000-2001 and 2002-2003. The difference between the 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 in terms of its incentives to enter ex-Red locations. In the first sub-period, Blue firm knew that Red firm would not re-enter at those ex-Red locations after the Red store closings, which removes the preemptive motives for Blue to enter those locations. On the other hand, in the second sub-period, there is no preemptive motive for Blue firm to enter at any potential location, including the ex-Red ones, since the expansion of Red firm has stopped. Thus, the ex-Red locations are not less favorable compared with the other potential locations, 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 the period between 2000 and 2001, when Red firm is still expanding: 5.13 million dollars per store. The termination of Red firm's expansion makes ex-Red locations less unattractive compared to other potential locations, with a median subsidy size of 1.29 million dollars per store. One explanation for the difference between the subsidies needed for Blue to enter in the two sub-periods is preemptive incentives. Preemption leads to the unattractiveness of the locations that a Red store used to occupy, compared with the other potential locations; therefore, many retail spaces stayed empty for years.

To better interpret the size of the subsidies, I compare them to the size of observed subsidies given to Blue firm between 2000 and 2014. The subsidy data came from goodjobsfirst.org. Although the list of subsidies is incomplete and some of the subsidy sizes are approximated, it gives a sense of the size of the subsidies in general. The average size is about 0.5 million dollars. Accordingly, I count the number of ex-Red locations whose payoffs of entry are less than half a million dollars lower than the payoff of the median potential location that neither firm has ever entered. 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 to four locations per period is observed between 2002 and 2003. Overall, the observed average size of 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 stores when many Red stores closed down. This is not intended to measure the total welfare change due to the store closings. Other factors such as employment, impact on small businesses, and local government income are also affected by exits of big box retail stores (Basker, 2007,Jia, 2008). However, the analysis helps to clarify whether, in general, the welfare loss is comparable to the size of subsidies. I use the demand model to compute the change in distance between a consumer and a store in the consumer's choice set due to each of Red store's closings. Note that some consumers may switch to an outside option after a Red store closes. For those consumers, I assume the distance traveled to the outside option to be the average distance traveled by a consumer to a discount retail store, 15 miles, which 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$\times$7.25(federal min. wage)$\times$10 trips$\times$2(round trip)/2.5(avg. household size).[25] The total welfare loss is about 1.26 million dollars per year. Although much

[25]The total distance divided by driving speed of 40 miles per hour is the total time of travel, which is multiplied by the federal minimum wage to get the dollar value. I assume that a consumer makes ten trips per year to a discount store. Given that the estimated annual spending is $2444, it seems reasonable to assume that each consumer spends about $240 on each trip. Ten trips per year is also much lower than the estimate by Fox *et al.* (2004) of each consumer visiting a discount store once every two weeks, on average, which leads to a lower bound for consumer welfare loss. Then, the result is multiplied by two to account for round trips and divided by the average household size from the census

smaller than the average subsidy of 3.21 million dollars, the total welfare loss for the 2002-2003 sub-period, 1.10 million dollars, is close to the 1.29 million dollar average subsidy needed for Blue firm to enter the market in the same sub-period. Therefore, consumer welfare loss due to Red store closings is substantial and is comparable to the size of the subsidy necessary to attract a new store.

7 Conclusion

This paper studies how multi-store retail chains make entry decisions, with an 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 and demand data. As a result, the spatial interdependence across store locations is preserved. The model is estimated by solving the dynamic game using backwards induction and applying a 'rolling window' approximation to compute the value function. The model and the estimation method can be applied to other retail industries or sectors in which network effects or economies of scale and scope 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 conducted to quantify preemptive incentives and evaluate local governments' subsidy policy. The results suggest that preemptive incentives are important in multi-store retailers' entry decisions and that they can lead to substantial loss of producer surplus. When a retailer exits a market, as frequently observed since the early 2000s, the store location becomes less attractive to other retailers due to the absence of preemptive incentives to entry. In these cases, although consumer welfare loss from store closings can be significant, average government subsidies prove insufficient to encourage entry by other retailers. The framework of the analyses can be used to study strategic interactions between companies in other industries and to assess public policy issues that arise in those markets.

data, assuming that one person shops for each household.

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 σ^*_{-j} and σ^*_j are separable, $\exists m \in \{1, ..., M\}$, s.t. $\sigma^*_{jm}(s_m, \sigma^*_{-jm}(s_m), B^*_m) \neq \sigma'^m_j(s, \sigma^*_{-jm}(s_m), B)$, and $\sum_{l\in P_m} \pi_{jl}(\sigma'^m_j(s), \sigma^*_{-jm}(s_m), s) > \sum_{l\in P_m} \pi_{jl}(\sigma^*_j(s_m), \sigma^*_{-jm}(s_m), s_m)$. But σ^*_{jm} is the best response of σ^*_{-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. $\sum_{l\in P_m} \pi_{jl}(\sigma'^m_j(s), \sigma^*_{-jm}(s_m), s) \leq \sum_{l\in P_m} \pi_{jl}(\sigma^*_j(s_m), \sigma^*_{-jm}(s_m), s_m)$.

Proof of Theorem2: If all conditions are satisfied, $\pi(s)$ is additively separably in $\{P_1, \cdots, P_M\}$. By results in Gorman (1959), $\{1, \cdots, L\}$ is separably in $\{P_1, \cdots, P_M\}$.

Definition A.1 *Locations $\{1, \cdots, L\}$ are <u>separable</u> in the partition $\{P_1, \cdots, P_M\}$ if*

$$\frac{\Delta\mathbb{E}V(s_j, s_{-j}, l)}{\Delta\mathbb{E}V(s_j, s_{-j}, h)} \perp (s^k_j, s^k_{-j}) | \{B_{jm}, B_{-jm}\}^M_{m=1}, \; \forall l, h \in P_m, \; \forall k \notin P_n, m \neq n,$$

where l, $h \in P_{m_l}$, and $k \in P_{m_k}$, and

$$\Delta\mathbb{E}V(s_j, s_{-j}, l) = \mathbb{E}V(s^l_j = 1, s^{-l}_j, s_{-j}) - \mathbb{E}V(s^l_j = 0, s^{-l}_j, s_{-j}).$$

Definition A.2 *Firm j's strategy σ^*_j is <u>separable</u> in the partition $\{P_1, \cdots, P_M\}$ if for given σ_{-j}, $\exists \sigma^*_{j1}, \sigma^*_{j2}, \cdots \sigma^*_{jM}$ s.t.*

$$\sigma^*_{jm}(s_{jm}, s_{-jm}, B_m) = \sigma^{*m}_j(s_j, s_{-j}, B),$$

where $\sigma^*_j = \left(\sigma^{*1}_j, \cdots, \sigma^{*M}_j\right)$, $B_m = (B^*_{jm}, B_{-jm})$, $B^*_{jm} = \sum_{l\in P_m} a^{*l}_j$, $B_{-jm} = \sum_{l\in P_m} a^l_{-j}$, $\forall m = 1, \cdots, M$, and $\sum^M_{m=1} B_m = B$.

Theorem A.1 *If locations $\{1, \cdots, L\}$ are separable in the partition $\{P_1, \cdots, P_M\}$, there exists a separable equilibrium.*

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

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

1. *$R^l_j(s)$ is additively separable in partition $\{P_1, \cdots, P_M\}$,*
2. *Distribution cost and fixed cost at location l is independent of z^k_j and x^k_j, where $k \in P_n$, $m \neq n$,*
3. *η^l_j are independently distributed across markets.*

PROOF OF THEOREM A.2: Prove by induction. Rewrite the value function as

$$\mathbb{E}V(s_{it}, s_{-jt}) = \sum_{\tau=t}^{\infty} \beta^{\tau-t} \left(\sum_{s_\tau} \mathbb{E}\pi(s_\tau) P(s_\tau | s_t) \right) = \sum_{\tau=t}^{\infty} \beta^{\tau-t} \mathbb{E}V_\tau(s_t).$$

The first term in the outer sum, $\mathbb{E}V_\tau(s_t) = \mathbb{E}\pi(s_t)$ when $\tau = t$, is separable in $\{P_1, \cdots, P_M\}$ by Theorem 2. Assume $\mathbb{E}V_\tau(s_t)$ is separable in $\{P_1, \cdots, P_M\}$ for $\tau = T$, then $\sum_{s_T} \mathbb{E}\pi(s_T) P(s_T|s_t)$ is separable. Apply two-stage budgeting, $P(s_T^l | s_t, \{B_{mt}\}_t^T)$ 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 $\mathbb{E}V_{T+1}(s_t)$ is separable.

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

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

Then

$$\Delta\mathbb{E}V_{T+1}(s_{jt}, s_{-jt}, l_{T+1})$$

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

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

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

where $s_{jmT} = \{s_{jT}^h | h \in P_m\}$, and $s_{-jmT} = \{s_{-jT}^h | h \in P_m\}$. Note

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

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

where $s_{jT+1} = s_{jT} + l_{T+1}$, and $s'_{-jT+1} \in \{s_{-jT+1} | s_{-jT+1} = s_{-jT} + h_{T+1}, h \in P_m\}$,

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

where $s_{jmT+1} = s_{jmT} + l_{T+1}$. Thus $\sum_{s_{-jT+1}} \left[\left[\mathbb{E}\pi(s_{jT} + l_{T+1}, s_{-jT+1}) - \mathbb{E}\pi(s_{jT}, s_{-jT})\right] P(s_{-jT+1}|s_T)\right]$ is additively separable in $\{P_1, \cdots, P_M\}$. Since $\mathbb{E}V_T(s_t)$ is separable,

$$\frac{\Delta\mathbb{E}V_{T+1}(s_{jt}, s_{-jt}, l_{T+1})}{\Delta\mathbb{E}V_{T+1}(s_{jt}, s_{-jt}, h_{T+1})} \perp (s_{jt}^k, s_{-jt}^k) | \{(B_{jmt}, B_{-jmt})_{t=1}^{T+1}\},$$

$\forall l, h \in P_m$, 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 $\{P_1^r, \cdots P_M^r\}$.

3. Given demand estimate $\hat{\beta}^r$ and market division $\{P_1^r, \cdots P_M^r\}$, 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.

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

	Stores and distribution centers in 2001	
	Blue	Red
Number of stores	2698	1883
Total number of distribution centers	35	18

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

	Blue	Red
distance to closest competitor's store	8.38	3.46
std. dev.	20.11	11.22
distance to closest same firm's store	11.90	10.16
std. dev.	15.12	20.79
total number of stores	1983	1140
number of same firm's stores in 30mi	4	3
number of any stores in 30mi	7	9
population density (10^5)	1.04	8.16
std. dev.(10^5)	3.20	1.86
distance to distribution center	98.47	126.56
std. dev.	71.41	122.02
total number of distribution centers	35	18

Table 3: Summary statistics of block group demographics.

	1980	1990	2000
mean population	0.83	1.11	1.35
mean income per capita	14.73	18.56	21.27
mean share African-American	0.10	0.13	0.13
mean share elderly	0.12	0.14	0.13
mean share young	0.35	0.31	0.31
no. of observations	269,738	222,764	206,960

source: U.S. census 1980, 1990, 2000

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

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

Table 6: Demand estimates

Average weekly sales in $1000s by store

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

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

Table 7: Demand comparative statics

Probabilities of shopping at Red store

	Population density			
Distance	1	10	50	250
2	0.18	0.14	0.06	0.01
4	0.34	0.21	0.07	0.02
10	0.80	0.33	0.08	0.02
20	0.88	0.34	0.08	0.02

Table 8: Clustering results comparison with CBSA markets

	Clustering	CBSA
Max loss per location $max_l\,[\lvert S(l)-S(l\in r^*)\rvert/S(l)]$	0.005	0.530
Total loss $[\sum_l \lvert S(l)-S(l\in r^*)\rvert]\,/\,[\sum_l S(l)]$	6×10^{-5}	8×10^{-3}
Affected locations $\sum_l \mathbb{1}_{S(l)\neq S(l\lvert l\in r^*)}$	55	123
Total locations	241	241

Table 9: Distribution and fixed cost estimates

Parameter estimates and 95%confidence intervals	
Blue firm's distribution cost ($1000/mi)	1.61
	[1.23, 1.98]
Red firm's distribution cost ($1000/mi)	0.68
	[0.02, 0.80]
Blue firm's fixed cost ($M)	0.43
	[0.27, 0.55]
Red firm's fixed cost ($M)	0.15
	[0.09, 0.31]
Number of observations	1226
likelihood ratio index	0.32

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

Choice probabilities and payoffs	
Average payoff increase($M)	0.86
Average choice probability decrease	0.14
Number of delayed entries	425
Total number of obs.	1278
Number of Blue stores out of 425	392
Efficiency loss ($M)	397

Table 11: Preemption vs. no preemption: location comparison

Characteristics of preemption and no preemption locations

	Preemption	No preemption
Distance to Blue DC (mi)	217.94	208.50
Distance to Red DC (mi)	273.22	287.35
Total Store density	24.62	22.56
Blue store density	12.14	11.55
Population density (1000)	175.48	172.12
Number of observations	425	853

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

Change of choice probabilities and payoff

Average payoff increase($M)	2.99
Probability of entry	0.77
Number of entries	472
Total number of obs.	1278

Table 13: Subsidies before and after Red firm stops expanding

Average subsidies per store

	Total	2000-2001	2002-2003
Median predicted subsidy ($M)	3.21	5.13	1.29
Number of subsidies $\leq$ 0.5M per period	5.5	3.5	7.5

Table 14: Consumer welfare loss due to store closings

Consumer drive time loss per store per year

	Total	2000-2001	2002-2003
Distance per person (mi)	4.05	4.10	3.99
Total distance (10^5mi)	8.70	9.80	7.60
Total welfare loss ($M)	1.26	1.42	1.10

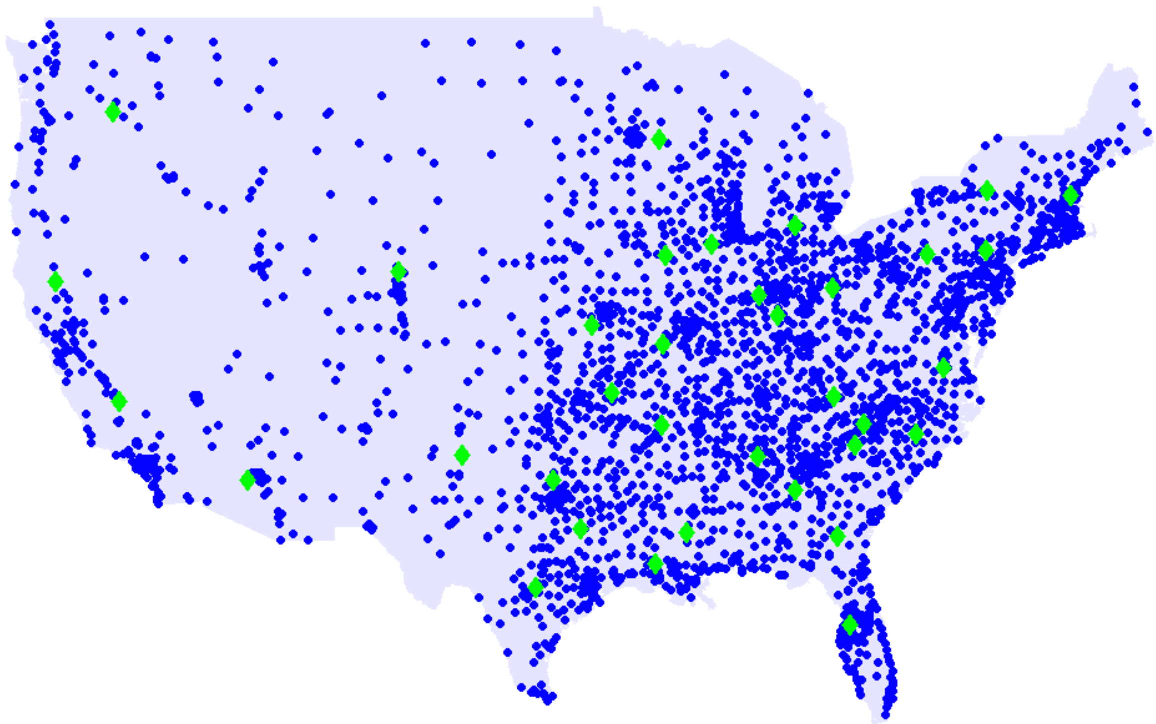

Figure 1: Blue stores and distribution centers in 2001

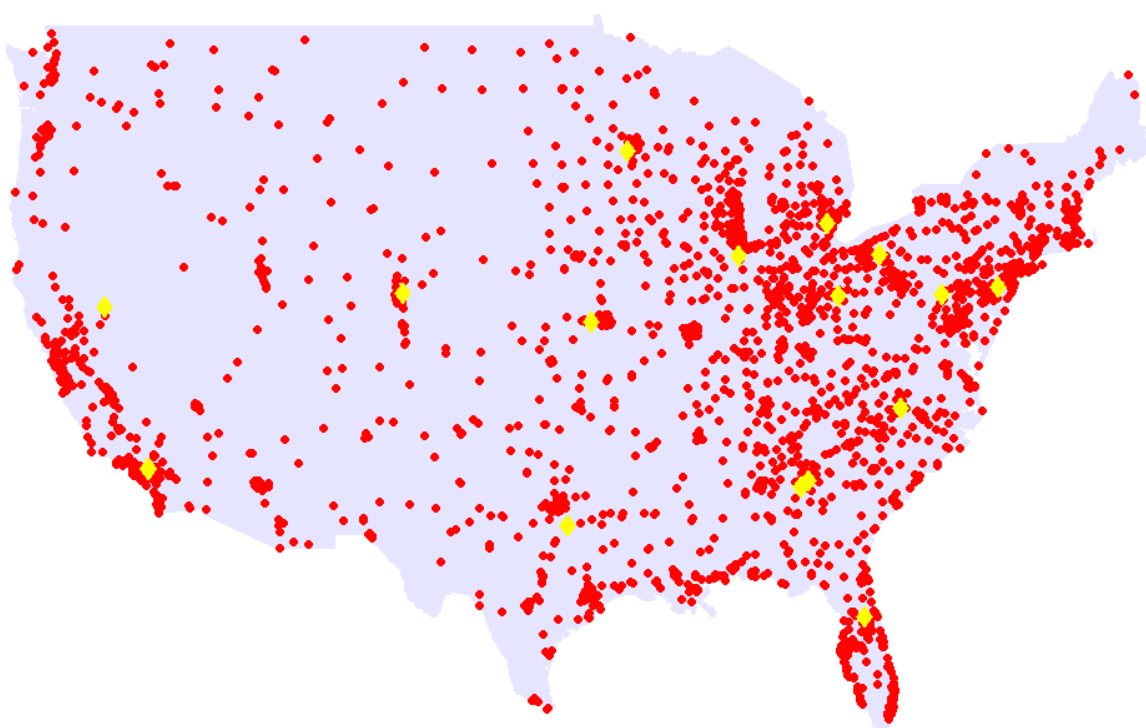

Figure 2: Red stores and distribution centers in 2001

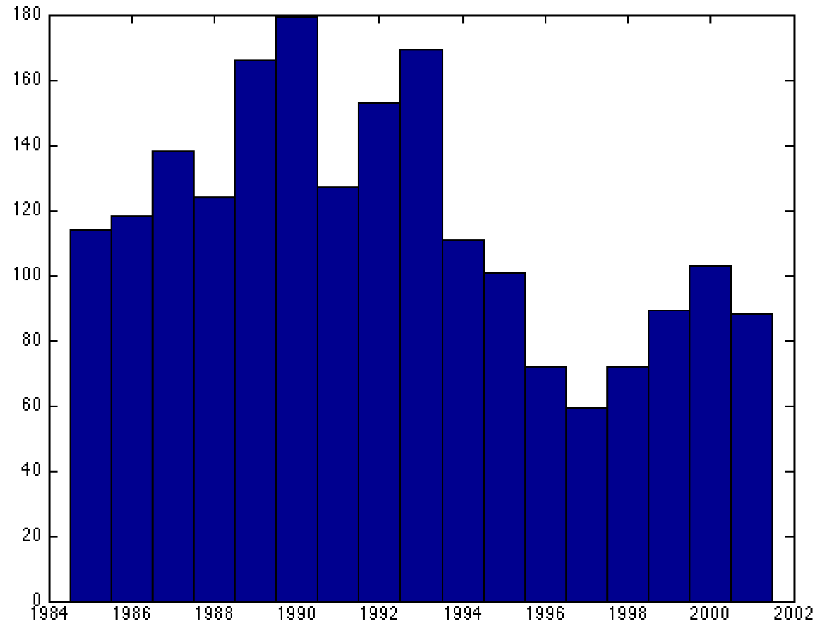

Figure 3: Blue store openings by year 1985-2001

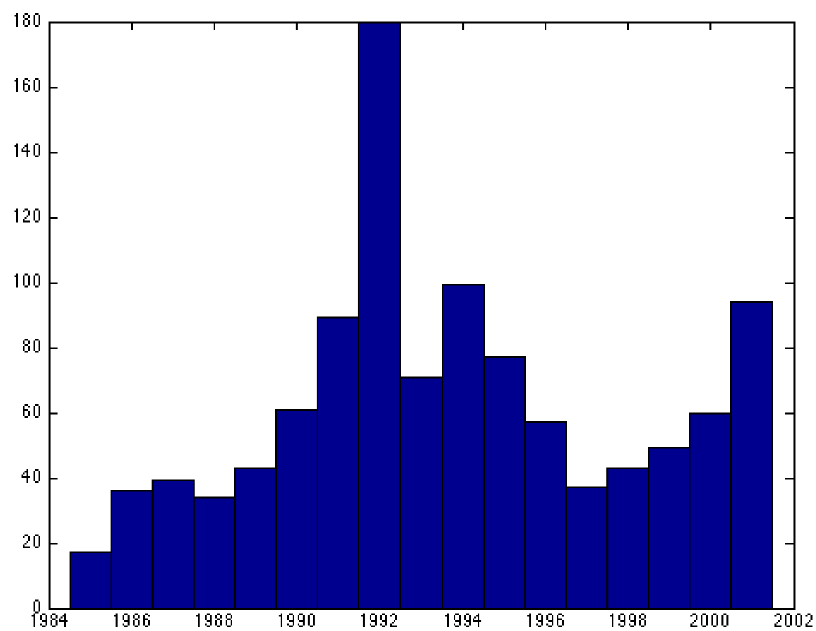

Figure 4: Red store openings by year 1985-2001

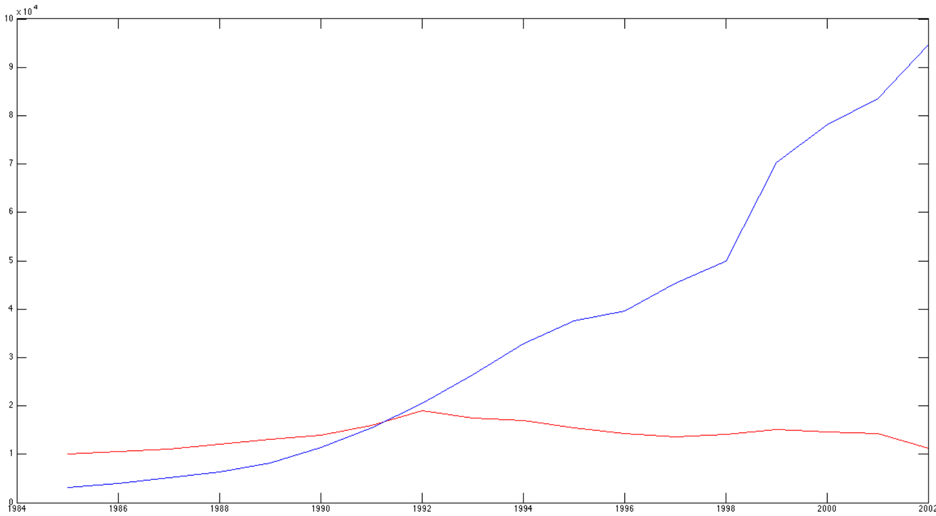

Figure 5: Book value of total assets, 1985-2002

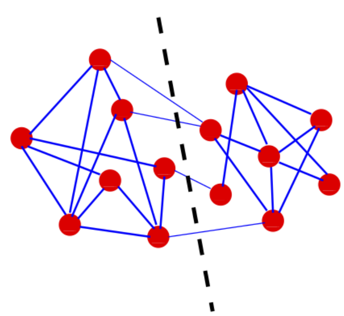

Figure 6: Graph partitioning[a]

[a]Fortunato and Castellano, 2009

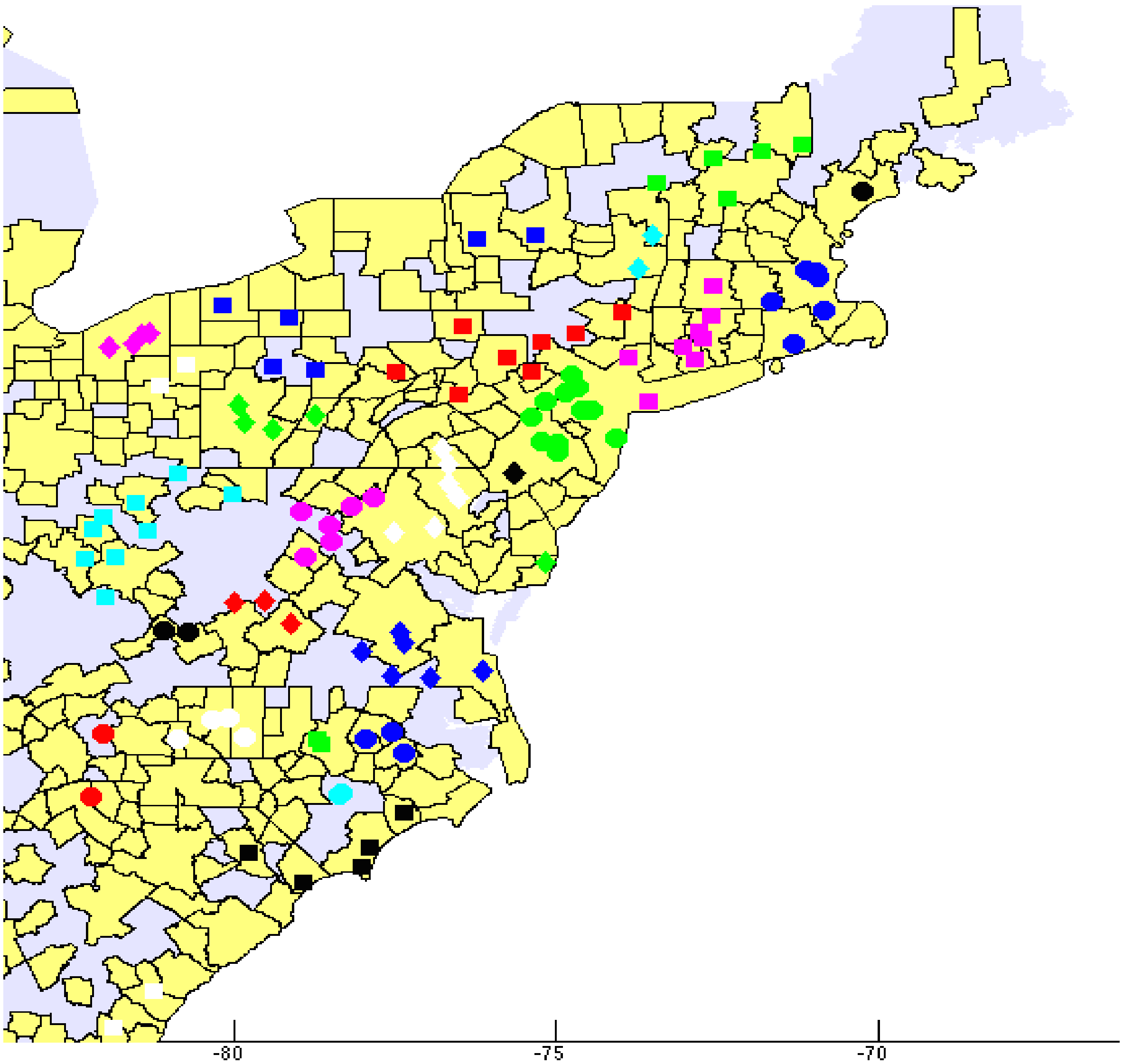

Figure 7: Markets by clustering and CBSAs, 1997Q3

References

Bajari, Patrick, Benkard, C Lanier, and Levin, Jonathan. 2007. Estimating Dynamic Models of Imperfect Competition. *Econometrica*, **75**(5), p. 1331–1370.

Basker, Emek. 2007. The Causes and Consequences of Wal-Mart's Growth. *The Journal of Economic Perspectives*, p. 177–198.

Berry, Steven. 1992. Estimation of a Model of Entry in the Airline Industry. *Econometrica*, p. 889–917.

Berry, Steven, Levinsohn, James, and Pakes, Ariel. 1995. Automobile Prices in Market Equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Bradley, Stephen, Ghemawat, Pankaj, and Foley, Sharon. 2002. *Wal-Mart Stores, Inc.*

Bresnahan, Timothy, and Reiss, Peter. 1991. Entry and Competition in Concentrated Markets. *Journal of Political Economy*, p. 977–1009.

Ellickson, Paul, and Misra, Sanjog. 2008. Supermarket Pricing Strategies. *Marketing Science*, **27**(5), p. 811–828.

Ellickson, Paul, Houghton, Stephanie, and Timmins, Christopher. 2013. Estimating Network Economies in Retail Chains: a Revealed Preference Approach. *The RAND Journal of Economics*, **44**(2), p. 169–193.

Fortunato, Santo, and Castellano, Claudio. 2012. Community Structure in Graphs. *Pages p. 490–512 of: Computational Complexity.* Springer.

Fox, Edward, Montgomery, Alan, and Lodish, Leonard. 2004. Consumer Shopping and Spending Across Retail Formats. *The Journal of Business*, **77**(S2), p. 25–60.

Gorman, William. 1959. Separable Utility and Aggregation. *Econometrica*, p. 469–481.

Gorman, William. 1971. Two-Stage Budgeting. *Unpublished Paper - L.S.E. Working Paper.*

Hausman, Jerry, and Leibtag, Ephraim. 2007. Consumer benefits from increased competition in shopping outlets: Measuring the effect of Wal-Mart. *Journal of Applied Econometrics*, **22**(7), 1157–1177.

Holmes, Thomas. 2011. The Diffusion of Wal-Mart and Economies of Density. *Econometrica*, **79**(1), p. 253–302.

Igami, Mitsuru, and Yang, Nathan. 2014. Cannibalization and Preemptive Entry in Heterogeneous Markets. *Unpublished Paper - Yale Working Paper.*

Jia, Panle. 2008. What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry. *Econometrica*, **76**(6), p. 1263–1316.

Mazzeo, Michael J. 2002. Product Choice and Oligopoly Market Structure. *The RAND Journal of Economics*, p. 221–242.

Neumark, David, Zhang, Junfu, and Ciccarella, Stephen. 2008. The Effects of Wal-Mart on Local Labor Markets. *Journal of Urban Economics*, **63**(2), 405–430.

Pakes, Ariel. 1986. Patents as Options: Some Estimates of the Value of Holding European Patent Stocks. *Econometrica*, **54**(4), 755–784.

Pakes, Ariel, Ostrovsky, Michael, and Berry, Steven. 2007. Simple Estimators for the Parameters of Discrete Dynamic Games (With Entry/Exit Examples). *The RAND Journal of Economics*, **38**(2), 373–399.

Reinganum, Jennifer. 1981. Market Structure and the Diffusion of New Technology. *The Bell Journal of Economics*, p. 618–624.

Rust, John. 1987. Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, p. 999–1033.

Ryan, Stephen P. 2012. The Costs of Environmental Regulation in a Concentrated Industry. *Econometrica*, **80**(3), 1019–1061.

Schmidt-Dengler, Philipp. 2006. The Timing of New Technology Adoption: The Case of MRI. *Unpublished Paper - L.S.E. Working Paper*.

Seim, Katja. 2006. An Empirical Model of Firm Entry With Endogenous Product-Type Choices. *The RAND Journal of Economics*, **37**(3), p. 619–640.

Shoag, Daniel, and Veuger, Stan. 2014. Shops and the City: Evidence on Local Externalities and Local Government Policy from Big Box Bankruptcies. *Unpublished Paper - Harvard Kennedy School Working Paper*.

Zhu, Ting, and Singh, Vishal. 2009. Spatial Competition with Endogenous Location Choices: An Application to Discount Retailing. *Quantitative Marketing and Economics*, **7**(1), p. 1–35.