

**Sleeping with the “Frenemy”:
The Agglomeration-Differentiation Tradeoff
in Spatial Location Choice[†]**

Sumon Datta

K. Sudhir

October, 2008

Preliminary Draft

Please do not cite or circulate

[†] Sumon Datta is a doctoral student and K. Sudhir is Professor of Marketing at the Yale School of Management. This is the third essay from Datta’s doctoral dissertation and his job market paper.

Abstract

A central tradeoff in location choice is the balance between agglomeration and differentiation. Should a firm co-locate (sleep) close to a competitor to increase volume (competitor is a “friend” who can draw more customers to the location with agglomeration) or locate far away from a competitor in order to reduce competition and increase margin (competitor is an “enemy” from who one should spatially differentiate)? Extant structural models ignore the agglomeration benefit, but given the high levels of co-location, this is a serious gap in the literature. Since observed co-location may be consistent with pure differentiation rationales such as (a) high demand at the location; (b) low cost at the location and (c) restrictive zoning regulations which allow entry in only small areas, it is challenging to disentangle the agglomeration-differentiation tradeoff from firm location choices.

The paper develops a comprehensive structural model of entry and location choice that helps disentangle the agglomeration-differentiation tradeoff. The paper has three key innovations: First, to separate the agglomeration-differentiation effect, we decompose profits into revenue and cost and then further show how to decompose the revenue into its components of consumer choice based volume and competition based price. Second, to separate demand and cost explanations, and to non-parametrically identify price competition, we augment entry and location data with revenue data for all stores and price data for one store chain in our dataset. Finally, to capture zoning effects, we introduce a new approach to obtain zoning data, an approach that should be of general interest for a large stream of spatial location applications

Our estimates and counterfactual analysis show that the agglomeration effect is strong and explains a significant fraction of observed co-location. Surprisingly, zoning has little direct effect on co-location. But tighter zoning restrictions interact with the agglomeration effect to explain a surprisingly large fraction of observed co-location. We find that a small change in zoning causes a discontinuous impact on the observed location pattern. This highlights the value of a structural model in understanding how a small perturbation of market characteristics can cause strategic firms to respond in complex and nonlinear ways.

Keywords: Retailing, Entry, Differentiation, Agglomeration, Zoning, Grocery, Empirical Industrial Organization

1. Introduction

Spatial clustering is a common phenomenon in many types of retail markets such as restaurants, automobile dealerships, electronics shops and bridal boutiques. That the phenomenon is well recognized in the popular imagination is seen in the popular labels for such retail clusters: hamburger alleys, restaurant rows, automobile malls etc. Such clustering is not always obvious or transparent. Consider for example, the retail locations of competing grocery stores. Figure 1 shows the distribution of distance between a grocery store and its nearest competitor in the three US states of New York, Pennsylvania and Ohio. Somewhat surprisingly, over 47% of stores are located within 0.5 miles of a competitor. What explains the high observed levels of co-location in grocery stores?

The Agglomeration-Differentiation Tradeoff

Economists have long recognized that the presence of a nearby competitor can increase profits if the co-location increases aggregate demand around the location even though it may suppress prices (Marshall, 1920). Agglomeration at a location may attract more consumers as it reduces search costs or travel costs when consumers search for products or prices (Stahl, 1982; Bester, 1998; Fischer and Harrington, 1996). Store co-location can also resolve consumers' hold-up problem; the problem of firms raising prices once the consumer arrives at the store because there is no nearby competitor (Wernerfelt, 1994). Note that the increase in demand due to co-location may be the result of stealing business from competitors in other locations. It does not require the market to expand in the traditional sense where non-users are converted to users or where usage rates are increased.²

There is also some empirical evidence of the benefits of agglomeration. Fox et al. (2007) find evidence of agglomeration benefits for a single market using household level data. Watson (2005) finds partial evidence that a single nearby rival could have a positive effect on profits, but his model does not offer any economic explanation for this observation. Vitorino (2007) finds evidence for inter-store spillovers in a particular kind of retail cluster - shopping malls. But in a mall setting, firms only make a strategic entry decision; they do not face the trade-off of whether to co-locate or spatially differentiate with rivals.

² For example, in the grocery market, *business stealing* is the plausible explanation for such agglomeration benefits whereas *market expansion* is unlikely.

Agglomeration can increase store traffic or volume by raising the likelihood of consumer visits and purchase.³ In this framework, a competing store is a “friend.” In contrast, most basic economic models emphasize the benefits of spatial differentiation. Here, retail stores seek to spatially differentiate from other competitors to minimize price competition. In this framework the competing store is an “enemy;” the closer the enemy, the lower the profits. Overall, given the two frameworks, a competing store is a “frenemy”: a friend (increases aggregate demand) who is simultaneously an enemy (suppresses prices).⁴

Should one “sleep with the frenemy” (i.e., co-locate)? The answer depends on balancing the gains in volume from locating close to a competitor with the gain in price from differentiation by locating farther away. This is the *agglomeration-differentiation tradeoff* or the *volume-price tradeoff*. It is important to note that the magnitude of the volume effect would depend on the level of market characteristics. For example, if consumers are in fact attracted to a location with multiple stores then the gains in volume for a firm, from co-location with a rival, would be greater if the location is surrounded by high population or high income as opposed to co-location in a low population or low income neighborhood. Similarly, when a cluster of rivals is in a different location from the firm’s location, the loss of volume to the cluster of rivals would be greater in a high population or high income neighborhood (‘greater number of valuable consumers’ equals ‘more to gain / lose’). A simple average ‘net effect’ of competitors across all locations and all markets cannot explain such variations. Thus it is necessary to disentangle the volume and price effects of competitors.

How can we measure the volume and price effects due to competitors? A retailer could use detailed household level data on consumer store choices across several markets that vary in market characteristics (like population and income) and the number of stores of different formats, and their locations, to estimate the benefit of agglomeration (volume effect due to co-location) through a household level model of store choice.⁵ Given these household level estimates, one can then solve for a competitive pricing equilibrium to identify the benefit of spatial differentiation (price effect due to differentiation). Such a method, however, tends to be

³ Also see Varian (1980), Arentze et al., (2005), Konishi (2005), Wolinsky (1983) and Dudey (1990).

⁴ “Frenemy” is a portmanteau of “friend” and “enemy” increasingly used to refer to a partner that is simultaneously a competitor. (Wikipedia)

⁵ For example, Fox et al., (2007) use data from a multi-outlet panel to study consumers’ shopping behavior and its impact on store revenues. However their data is from a single major metropolitan market.

impractical because such detailed household level data across multiple retailers are difficult to obtain and a household level analysis across markets is too onerous.

Another approach could be to use firm level data on revenues and prices of all stores across several markets. Assuming store locations as given, one could develop a consumer shopping behavior model to identify the benefit of agglomeration, coupled with a price competition model to identify the benefit of differentiation. This approach, however, could suffer from serious issues due to the endogeneity of market structure (i.e., number of firms that enter a market and their locations). For instance, a location with a high unobserved demand shock is likely to have higher revenues but the location is also likely to attract more firms. Due to endogenous market structure, the bias in the parameters capturing the agglomeration benefit and the competitive interactions can also have important policy implications.

To infer the strategic interactions between firms, researchers in marketing and economics have adopted an alternative empirical approach that uses readily observed entry and location decisions of firms. The approach is built on the idea that firms take into account their competitors' actions when making their decisions. Thus, by solving for the location choice game between firms we can infer the strategic interactions between firms. A vast majority of papers have used this approach to study firms' entry strategies (e.g., Bresnahan and Reiss 1991; Mazzeo 2002; Vitorino 2008).⁶ A smaller literature has analyzed the strategic location choice decisions (e.g., Seim 2006; Watson 2005; Zhu and Singh 2007), where firms not only decide whether to enter into a market but if they enter, where to locate and how far to locate from a competitor.⁷

The structural models of location choice use a reduced form profit function that allows latent profit in a location to depend on the number of competitors, and their distances from the location. This reduced form approach is not sufficient to separate the agglomeration-differentiation tradeoff. Firstly, profit is revenue minus the cost at the location, and competitors affect profit through revenue. Thus, not separating revenue and cost can create problems in inference. For instance, even if the competition effect is high at a location, firms might still choose such a location due to low costs. Hence, as a first step, to isolate the effect of competitors, we decompose the reduced form profit model into revenue and cost. To this effect, we augment

⁶ Some have addressed entry decisions of retail chains, considering how these chains build up their network (Jia 2008; Holmes 2008; Ellickson, Houghton, and Timmins 2008).

⁷ A related literature models consumer choice of spatial locations but treats outlet locations as exogenous (Fox et al., 2007; Duan and Mela 2007; Thomadsen 2007). Chan, Padmanabhan, and Seetharaman (2006) also model consumer choice of gasoline outlets when a social planner determines the locations of those outlets.

the entry and location choice data with store revenue data. Secondly, we want to separate the ‘net effect’ of competitors into its components of consumer store choices (volume effect of competitors) and price suppression (price effect of competitors). Hence, we also decompose the revenue into its components of volume and price.

Since the location of competitors affects both the volume and the price, and consequently the potential revenue in a location, complete non-parametric identification of the two effects would ideally require additional data on sales or prices at all stores. Without this data, one would have to rely on suitable functional form assumptions so that the locations of rivals affect store volume differently from the way they affect store prices. Currently, we have price data for a set of stores that belong to one store chain. This price data aids the non-parametric estimation of the price effect of competitors while the revenue data for all stores helps identify the agglomeration benefits. Specifically, for the price effect, we follow a similar approach as in Seim (2006) and allow prices to be affected by the number of rivals, their types, and their distances from a location. For the volume effect, we model consumers’ shopping location choice behavior, which incorporates the number of firms and their geographic locations around consumers. Hence, we exploit two kinds of variations in the data - variation in the number and locations of stores around consumers which drives the volume at each store, and variation in the number and distances of rivals around a store which drives the prices at the store.

Alternative Reasons for Co-location

While the benefits of agglomeration can lead to store co-location, it is important to recognize that agglomeration effects are *not necessary* for stores to co-locate. Even if there are no agglomeration benefits, firms may locate close to each other for at least four reasons. First, if demand is extremely high in a location relative to neighboring locations, firms may choose to co-exist and share the high demand rather than differentiate. Second, if the cost in a location is extremely low, then again firms may co-locate in the low cost location. Third, competing stores may locate close to each other because of zoning regulations, which restrict the scope of spatial differentiation. Finally, if stores differentiate through their formats, the need for spatial differentiation may be lower.

Retail format is an important moderating factor in entry and location choice. Firms can differentiate from rivals through their store format if different formats specialize in different

product categories or pricing strategies (EDLP, Hi-Lo, etc). More importantly, if there are agglomeration benefits, such benefits may be greater when stores of different formats co-locate (*format agglomeration*) because then it allows consumers to get greater variety of products or prices in one location (Bester 1998; Fischer and Harrington 1996).

In this paper, we estimate a structural model of firm entry and location choice that isolates the four factors that affect the location decision: 1) the extent to which consumers value store agglomeration and format agglomeration; 2) demand; 3) cost; and 4) spatial retail zoning regulations. We also allow for the price competition effect to differ by retail formats.

Demand and Cost

We decompose the reduced form profit model into revenue and cost and we augment our entry and location choice data with store revenue data. We include demand factors in a consumer choice model. Observed cost factors, if any, affect the profit threshold for firm entry and location choice. Extant models that treat profits in reduced form may not have such a structural interpretation for demand and cost variables.

The challenge lies in dealing with unobserved (to the researcher) demand and cost factors. Demand shifters such as traffic patterns and store features are typically unobserved by the researcher. Also, cost shifters such as tax-breaks and store and location-specific cost variables may be hard to obtain. However, these features may be common knowledge to firms. Extant papers using only entry and location data group all unobserved demand and cost factors into a single profit shock at the location for the firm. Hence they cannot explain whether low cost or positive demand shock causes firms to co-locate at a location. This creates problems in inference. For example, if there is no store in a location with high population and income, because of unobserved high costs, existing models would downward bias the effects of population and income on profits.⁸

As discussed above, the store revenue data aids the identification of the agglomeration benefits. The portion of the revenue that is not explained by the observed demand factors is attributed to a demand shock (due to unobserved demand factors) which is a draw from a

⁸ Some recent researches that also make use of the information in post-choice outcome data include Draganska et al., (2007) and Ellickson and Misra (2007).

distribution. Using this distribution we control for all demand factors at a location.⁹ Finally, using observed firms' actions, and allowing for correlations between demand and cost, we identify the distribution of cost through latent profits.

Retail Zoning Regulations

Zoning is an omnipresent factor in retail location choice. Most locations with high populations and incomes are typically zoned as residential land and therefore do not have retail stores. If one does not control for zoning, as in extant models, one would downward bias the impact of income and population on profits to explain the absence of stores in such locations. Apart from restricting the location where firms can enter, zoning regulations have other implication for a retailer's decision on whether to enter a market and if so, where to locate. Restrictive zoning, with smaller and more concentrated retail zones, that force sellers to locate closer, lowers consumer search costs (due to easier comparison shopping), thereby increasing price competition (Ridley et al., 2007).¹⁰ In response to such increased competition and potentially lower profits, fewer firms may enter a market (Suzuki, 2008).¹¹

Even though the critical importance of spatial zoning is fairly well-known, extant research has completely ignored this issue because of lack of availability of zoning data on a national scale across many markets. Zoning regulations are made on a local, town-by town, basis and it is virtually impossible to construct a dataset of zoning regulations across any meaningfully large region covering several towns and cities in the US. Some researchers in urban planning have looked at spatial zoning issues within a single city but never across multiple markets.¹²

Given the significance of zoning in location choice, we introduce a new approach to obtain zoning data across markets on a national scale. We use a new dataset called National Land Cover Dataset (NLCD) that identifies commercial zones. In conjunction with Geographic Information System (GIS) tools such as *ArcGIS* and *Google Earth*, we identify retail zones that allow entry of large retailers. Digital land-cover data are among the most popular data sources in

⁹ There is a potential selection bias because we only observe revenue data for locations that were chosen. We are currently trying to address this selection problem.

¹⁰ Ridley et al., (2007) collect zoning ordinance data from 15 municipalities.

¹¹ Suzuki (2008) does not have spatial zoning data. He uses several indices that proxy for the stringency of land use regulations.

¹² For example, see Shlay and Rossi (1981); Ridley et al., (2007) is an exception. However, they collect data for only a very small number of markets

ecology and forestry research, but have been unknown in the marketing and economics literature. Ours is the first application of these data to identify retail zoning regulations.

Modeling and Estimation Approach

As discussed above, estimating the strategic interactions between firms through store revenue and price data and by assuming store locations as given can have serious problems due to the endogeneity of market structure. Hence, we develop a structural model of entry and location choice that incorporates the effect of competitors on prices and revenues. Like most empirical studies of market structure, we solve a static, simultaneous move, discrete choice game.¹³ Similar to Seim (2006), we employ an incomplete information set-up. Apart from the cost which is common knowledge to firms (as discussed above) we assume an additional idiosyncratic difference in costs among firms, which competitors know only in distribution.

We note the key differences with respect to extant models. As discussed earlier, we disentangle the agglomeration-differentiation tradeoff by first decomposing the reduced form profit model into revenue and cost and then further decomposing the revenue into its components of volume and price. The consumer model explicitly models how travel costs, store and format agglomeration affect consumers' shopping location decision. Here we allow for consumer heterogeneity in preferences for store locations and store formats. Competitive price effects vary by the extent of spatial and format differentiation. We adapt our model for zoning by allowing firms to enter and choose locations only within retail zones. There may be multiple zones within each market. Given the revenue and price data, the model specification includes separate revenue, price and cost shocks.

The estimation relies on identification of equilibrium format and location choice probabilities. In multi-agent strategic games, a critical challenge relates to the possibility of multiple equilibria in the model.¹⁴ Specifically, for a given vector of model parameters θ , if there exists more than one set of equilibrium probabilities of firms' actions, the likelihood function cannot be well defined in a nested fixed-point (NFXP) algorithm. One way to deal with this

¹³ Since we do not have entry dates of stores, we are not able to solve a dynamic choice game. However, our model can be extended to a dynamic set-up.

¹⁴ See Aguirregabiria et. al., (2008) for a discussion on the distinction between multiple equilibria in model and multiple equilibria in data. They also discuss the various estimation approaches for discrete choice games with strategic interactions.

problem is to provide sufficient conditions for uniqueness of equilibria that the model parameters, θ , must satisfy (e.g., Siem, 2006 and Zhu and Singh, 2007).

Researchers have developed two-step estimation approaches that reduce the computational costs of NFXP and avoids the pitfalls of multiple equilibria. (e.g., Bajari et al., 2007a). In a two-step Pseudo Maximum Likelihood (PML) approach, the conditional choice probabilities are estimated in a parametric or nonparametric first stage and the model parameters are obtained by maximizing the resulting pseudo likelihood function.¹⁵ However, in most empirical contexts, consistent and precise first-stage estimates are infeasible and it is difficult to incorporate common unobservables in the first stage. A recursive extension of the two-step PML, called the Nested Pseudo Likelihood (NPL) approach addresses these drawbacks at a relatively small additional computational cost (Aguirregabiria and Mira, 2007). NPL requires an initial guess of the conditional choice probabilities which need not be consistent. The focus here is on the estimates found in the limit (called the NPL fixed point).

The PML and the NPL approaches assume that only one equilibrium is being played in the data. Su and Judd (2007) developed a Mathematical Programming with Equilibrium Constraints (MPEC) approach that avoids any assumption about the equilibrium being played.¹⁶ Vitorino (2007) is an application of this approach in the context of an entry model. Since the MPEC approach simultaneously estimates model parameters and conditional choice probabilities, its performance for models with large numbers of decision choices and state spaces (as in our location choice model with heterogeneous formats) is unclear.¹⁷ We therefore estimate the entry and location choice model using the NPL.

Typically the literature on two-step approaches, including NPL, considers fully dynamic games. We adapt the two-step NPL approach for our static game. This adaptation is similar to that of Bajari et al. (2006) who adapt their two-step approach to estimate a static game of strategic interactions.

¹⁵ Alternatives to likelihood based approaches include method of moments (Thomadsen 2005; Draganska et al., 2007; Pakes et al., 2007), minimum distance or asymptotic least square estimators (Pesendorfer and Schmidt-Dengler 2003; Bajari et al., 2006) and maximum score estimators (Bajari and Fox 2006; Fox 2007; Ellickson et al., 2008).

¹⁶ Other ways of dealing with multiple equilibria can be found in Bajari, et al., (2007b) and Ciliberto and Tamer (2006).

¹⁷ We were able to estimate a simpler version of our location choice model with homogeneous formats using the MPEC approach.

The rest of this article is organized as follows: Section 2 describes the model and estimation strategy. Section 3 describes the data and the data creation approach for zoning regulations. Section 4 describes the estimates of the model. Section 5 presents the results of counterfactual simulations. Section 6 concludes.

2. Model

Modeling firms' location choice decisions in continuous geographical space is practically impossible. Hence, we discretize the geographical space. For purposes of easy visualization, we illustrate using a square city, which is discretized into a grid of, say, L square blocks or 'locations' (See **Figure 2a**). Due to zoning regulations, only a subset of these locations is available for entry. Suppose there are Z sets of contiguous locations available for entry that we refer to as retail zones. In the illustrative picture, there are two such sets of contiguous locations (i.e., $Z = 2$). Thus, each retail zone has several locations where firms are allowed to open retail stores. That is, these are the '*potential retail locations*' (See **Figure 2b**). Consumers, however, may be present in all locations of the city and can choose which retail location to visit for shopping.

Firms, make entry and location choice decisions such that they maximize their profits in equilibrium. The profitability of choosing a particular location depends on firm revenue and the cost at that location. Because we want to distinguish between agglomeration benefits that arise from consumer shopping behavior and the differentiation benefits that result from competition moderation, we decompose firm revenue into its structural components: revenue potential of a location based on consumers' shopping location choice (allowing for consumer utility from endogenous agglomeration of stores) and price suppression due to competitive effects.

The model set-up proceeds in the following steps. We first model consumers' shopping location choice which is based on consumers' preferences for store agglomeration, store formats and retail location characteristics. From this we derive a dollar value for the potential revenue that a store with a particular format could make at a location if there was no competition effect. We then model the competition effect due to rival firms, which erodes this revenue potential through price suppression. This gives us a prediction of firm revenues at a location. Next, we account for demand and cost factors that are common knowledge to firms but not observed by

the researcher. Finally, we incorporate firms' private information about cost at a location to get the profitability at the location.

2.1 Consumers' Shopping Location Choice

As we discuss later, in the data section, we do not have detailed information about consumers beyond the Census Block Group (CBG) level. Hence we use demographic data at the CBG level and assume that consumers are located at the population density weighted center of their respective CBG.¹⁸ We allow consumers to shop at any of the grocery stores within a radius, R .¹⁹ We use this flexible set-up so that consumers are allowed to shop in a grocery store which may be in a neighboring market. Note that this automatically implies that the *trade radius* of a store (region from which a store gets its customers) is R .

Consumers incur a travel cost which could be non-linear with respect to the distance they have to travel. A non-linear specification allows us to test for decreasing marginal effects of distance. To allow for heterogeneity in travel costs we allow it to differ by the median household income of the CBG (med_hhI) and the median age (med_age). Zoning introduces another dimension of consumer heterogeneity. Because some locations may be purely residential locations, consumers in these locations have to travel a minimum distance (min_d) before they can get to the nearest retail location. Consequently, a consumer who is located far from the nearest retail location may be more willing to go to a store that is farther away, than say, a consumer who is close to several retail locations. For instance, a consumer who has to any way travel, say, 3 miles to get to the nearest retail location, may be less sensitive to travel to a store that is 4 miles away than a consumer who is surrounded by several retail locations within, say, a 2 mile radius. Hence, we allow the travel cost to also depend on the minimum distance to the nearest retail location. Formally, the cost of travelling to a retail location ' l ', for a consumer who is in CBG ' i ', is given by;

$$T_{il} = (\alpha_1 d_{il} + \alpha_2 d_{il}^2 + (\alpha_3 med_hhI_i + \alpha_4 med_age_i + \alpha_5 min_d_i) * d_{il}) \quad (1)$$

where, d_{il} is the distance the consumer must travel.

¹⁸ This is a standard assumption in the location choice literature.

¹⁹ Consumers are unlikely to travel very long distances for their grocery shopping. We estimate our model for $R = 6$ mi. and 8mi. The model estimates for the two specifications were substantively similar. Of course, R can take higher values but at the cost computational time.

A consumer may be willing to incur the travel cost if the store in that location has low prices, h_{kl} .²⁰ Also, there could be economies of scope from multipurpose shopping (also called one-stop shopping) when the retail location is closer to other retail businesses that cater to the consumer's non-grocery needs (e.g., electronics and apparel stores). To capture such scope economies of shopping, we consider the total retail business density in all retail locations within 2 mi. of location l , weighted by the distance from location l .²¹ Again, we allow business density to have a non-linear effect on consumers' utility. Next, because a consumer likely frequents locations where stores co-locate, we consider the effect of the presence of rival grocery stores in that particular location (i.e., *store agglomeration effect*). Furthermore, to capture scope economies of shopping within the grocery sector, we consider the effect of the presence of rival grocery stores of different formats in that particular location (i.e., *format agglomeration effect*). Finally, we allow for heterogeneity in consumers' intrinsic preference for different store formats, k ($k = 1, \dots, K$), through the intercept term.²²

Formally, for a consumer in CBG, i , the utility of shopping at a store of format k in retail location l is:

$$U_{ikl} = \lambda_{ik} - T_{il} - \gamma h_{kl} + \alpha_6 bus_l + \alpha_7 bus_l^2 + \alpha_8 N_l + \alpha_9 I_l^{MF} + \omega_{il} \quad (2)$$

where, λ_{ik} is consumers' intrinsic preference for store format k and is given by

$$\lambda_{ik} = \lambda_{0k} + \lambda_{1k} med_hhI_i + \lambda_{2k} med_age_i + \lambda_{3k} min_d_i \quad (3)$$

T_{il} Travel costs as described above

h_{kl} Price index of store format k at location l ;

bus_l distance weighted business density at location l ;

$(\alpha_6 bus_l + \alpha_7 bus_l^2)$ economies-of-scope due to non-competing retail businesses;

N_l number of competing grocery stores at location l ;

²⁰ h_i^k is the price index of a store of format k in location l . Its derivation is shown in sub-section 2.3. We use sales-weighted prices across all categories in a store as the price index of the store.

²¹ We tried other specifications such as just the business density at location l , and total businesses in a 2 mi. radius of location l (not weighted by distance) but found this specification to have the best explanatory power.

²² To the extent that co-location reduces consumers' search costs, we considered the effect of only those rivals who are in the same location. However, lower prices due to the competitive effect of all rivals (modeled in sub-section 2.3) might also affect the attractiveness of the location for consumers. We are currently modifying the consumer utility specification to include these price effects.

$(\alpha_8 N_l)$	store agglomeration utility;
I_l^{MF}	indicator for the presence of multiple store formats at location l ;
$\alpha_9 I_l^{MF}$	format agglomeration utility; and
ω_{il}	unobserved characteristics of location l (e.g., parking lots)

Assuming i.i.d. Type 1 extreme value distribution for the unobserved errors, and normalizing the intrinsic preference for one of the store formats to zero, the probability that a consumer at CBG i will visit a store with format k in retail location, l , is given by:

$$p_{ikl}^c = \frac{\lambda_{ik} - T_{il} - \gamma h_{kl} + \alpha_6 ret_l + \alpha_7 ret_l^2 + \alpha_8 N_l + \alpha_9 I_l^{MF}}{\sum_{k'=1}^K \sum_{j \in L_i^c} \lambda_{ik'} - T_{ij} - \gamma h_{k'j} + \alpha_6 ret_j + \alpha_7 ret_j^2 + \alpha_8 N_j + \alpha_9 I_j^{MF}} \quad (4)$$

where, L_i^c is the consideration set of a consumer at CBG, i , or the set of retail locations that lie within radius R of CBG i . Equation (4) is also the share of consumers in CBG i , who will shop in a store of format k at retail location l .

2.2 Linking Consumer Shopping Location Choice to Locations' Revenue Potential

To arrive at the revenue potential of location l for store format k , we aggregate the contribution of consumers of all CBGs who will shop in a store of format k at location l . Note that according to our specification, each location l can attract consumers from all CBGs that are within a radius, R (*trade radius*). We first specify the 'grocery spending potential' of consumers from CBG i in a store of format k at location l , (GSP_{ikl}). In particular, we weigh the location choice probabilities of consumers, derived from the utility model, by the number of such consumers and their spending potential, as represented by their per capita income²³:

$$GSP_{ikl} = p_{ikl}^c Pop_i PCI_i, \quad (5)$$

where

Pop_i	population of CBG i ;
$p_{ikl}^c Pop_i$	number of consumers of CBG i who visit a store of format k in location l ;
PCI_i	per capita income of consumers at CBG i .

²³ Instead of per capita income, one could use total per capita expenditure or per capita expenditure on grocery

To obtain the total grocery spending potential in a store of format k at retail location l , we sum the contributions from all CBGs within the trade radius R of location l :

$$GSP_{kl} = \sum_{i \in L'_l} GSP_{ikl}. \quad (6)$$

where, L'_l is the set of CBGs that lie within the trade radius of location l . Note that if consumers in a CBG are less likely to visit a particular store format in location l , then their contribution will be proportionately lower.

If we denote the vector of market characteristics of retail businesses, population, per capita income, etc. by X , the revenue potential for store format k at location l , which we denote as $f(X, N_l, I_l^{MF}; \alpha, \lambda, \delta)$, then can be specified as

$$f(X, N_l, I_l^{MF}; \alpha, \lambda, \delta) = GSP_{kl}^{\delta^k} \quad (7)$$

Thus, our specification results in a dollar value for the revenue potential of a store format at a location, which is a function of consumer preferences and is endogenous to firms' actions through the agglomeration benefits.

2.3 Competitive Effect

If a firm f with format k chooses location l , it must share the profit potential at location l with its rivals because the competitive effect would suppress price. We model this competitive effect in a flexible, semi-parametric manner so that we can split the profit potential differentially across firms as a function of their format and distance. For distance, we follow Seim (2006) and model competition in terms of discrete circular distance bands ($b = 1, \dots, B$) from the location l , up to its *trade radius*, R .²⁴ We allow for asymmetric competitive effects between store formats, such that the intraformat competitive effect may be different than the interformat competitive effect. Thus, two stores of different formats may compete less with each other than would two stores of the same format. Furthermore, the intraformat competitive effect may differ for different formats.

Let N_{kl}^b denote the number of rivals of format k in distance band b around location l . If we denote the price index as $h(N_k, \dots, N_{k'}, \dots, N_K; \beta)$, where N_t is the vector of number of rivals of format t in the different distance bands, the price index for firm f can be specified as:

²⁴ In our application, we have used R as the trade radius; the

$$h(N_k, \dots, N_{k'}, \dots, N_K; \beta, \tau) = \tau_k \prod_b (1 + N_{kl}^b)^{\beta_{kb}} \prod_b \prod_{k' \neq k} (1 + N_{k'l}^b)^{\beta_{k'b}} \zeta_{fl} \quad (8)$$

where, τ_k is a format specific parameter which allows the intrinsic pricing ability of firms to differ in their format (i.e., due to either cost or efficiency differences). The second component on the right-hand side is the intraformat competitive effect and the last component is the interformat competitive effect. Finally, ζ_{fl} is any unobserved price shock that the firm gets at location l (The h_{kl} term used in the consumer model is the price index less this price shock). We use a multiplicative specification to test whether the competitive effect of an additional rival decreases with the number of rivals. Adding 1 to the number of rivals captures the idea that in the absence of rivals within a particular distance band b , there is no competitive effect for firm f in that distance band.

Using price index data for all stores, one can non-parametrically identify the competitive effect of rivals. Currently, however, we only have data on prices for one store chain that exists in most of the markets in our dataset. Hence, for our identification, we partly rely on our functional form assumption of how rivals affect volume in a different way than how they affect prices.

2.4 Firm Revenue

The revenue that firms can get at a location depends on the total revenue potential, $f(\cdot)$, of a location and the competition effect, $h(\cdot)$. Whereas the benefits of agglomeration occur through increased revenue potential, the benefits of differentiation emerge from acquiring a greater share of that potential as a result of decreased competition.

$$\hat{R}_{fkl} = f(X, N_l, I_l^{MF}; \alpha, \lambda, \delta) * h(N_k, \dots, N_{k'}, \dots, N_K; \beta, \tau) \quad (9)$$

Now, there could be unobserved demand characteristics for a store at a retail location which might make the location particularly attractive for a firm. Hence, we assume that a store at a location gets a revenue shock which is firm- and location-specific and which is a draw from a log-normal distribution. The value of this revenue shock is common knowledge to firms but not known to the researcher. Adding this revenue shock enables us to scale our predicted revenue such that it matches with the actual observed revenue for the stores in locations that were chosen. Formally, the revenue of a store of format k at location l is given by:

$$R_{fkl} = f(X, N_l, I_l^{MF}; \alpha, \lambda, \delta) * h(N_k, \dots, N_{k'}, \dots, N_K; \beta, \tau) * (v_{fl})^\rho \quad (10)$$

where, v_{fl} is a draw from a standard log-normal distribution and ρ is the revenue parameter which is identified from the revenue data of stores at the locations that were chosen.

2.5 Firm Profits

Firms make their entry and location choice decisions based on the profitability of a location. Profits depend on the revenue that a firm can make at a location and the cost at the location. Good instruments for store and location-specific costs are usually hard to come by. However, certain cost factors may be common across several locations. Going back to our illustrative example, we had two sets of contiguous retail locations in the market. That is, we had two distinct ‘*retail zones*’. Now, retail zone specific cost factors that are common to all firms in the zone (e.g., zone tax breaks) may not be observed by the researcher, although they would be common knowledge for firms. Such retail zone specific cost factors might result in greater entry in one retail zone over another, even though they both are in the same market. To control for such endogenous entry of firms into retail zones, we allow for a retail zone specific cost factor.²⁵

Beyond the retail zone level, costs could vary across locations within each retail zone. We do not have any good instruments for location level costs in our data.²⁶ However, if we assume that the cost is correlated with the revenue shock at a location, we can infer this correlation from the revenue shock at the locations that were chosen and the number of firms that chose those locations. For example, if few firms choose locations with high revenue shocks, we can infer that such locations also have high costs. Consequently, instead of ignoring location level cost factors entirely, we assume that each location gets a cost shock that is proportional to the revenue shock. Formally, the profitability of a location is given by:

$$\Pi_{fkl}^z = R_{fkl} * \xi^z * (v_{fl})^{-c} \quad (11)$$

where, ξ^z is the retail zone specific scaling factor which controls for the unobserved attractiveness of the retail zone. It partly determines the number of firms that a retail zone can

²⁵ Traditionally, markets have been defined as counties or MSAs or towns and cities, and firms are assumed to make entry decisions into these markets. To control for endogenous entry of firms, due to market characteristics not observed by the researcher, these models incorporate a market specific effect in these models. However, there could be multiple retail zones within a market. When the researcher does not observe retail zone specific features, a homogeneous market specific effect would not sufficiently control for an endogenous entry into a particular retail zone. Hence, we deviate from the traditional approach of modeling firms’ entry decisions into markets and instead model entry decisions into specific retail zones and allow for unobserved retail zone specific effects.

²⁶ We tried using median property value in the nearest CBG as a cost variable but it turned out to have no explanatory power.

support. ‘ c ’ is the cost parameter which is identified from the revenue shocks and firms’ actions in the locations that were chosen.²⁷

Finally, we allow firms’ to have an idiosyncratic cost at a location. Formally, if a firm f , with format k , from an exogenous potential pool of F entrants ($f = 1, \dots, F$) decides to enter retail zone z and to choose location l then its profit is

$$\Pi_{fkl}^z = f(X, N_l, I_l^{MF}; \alpha, \lambda, \delta) * h(N_k, \dots, N_{k'}, \dots, N_K; \beta, \tau) * \xi^z * (v_{fl})^{\rho-c} * \eta_{fl} \quad (12)$$

Where, η_{fl} is an idiosyncratic firm- and location-specific cost factor. While the revenue and cost shocks are common knowledge for firms, we assume that η_{fl} is known only to the firm, and rivals know it only in distribution. Therefore, we introduce information asymmetry into the model, which we exploit for computing the equilibrium.

The proposed model nests simpler models with no agglomeration effects or those with homogeneous stores. Without agglomeration effects, the profit potential of a location is driven completely by exogenous demand characteristics X , as has been assumed in previous literature. Thus, the effect of the number of neighboring stores N_l and the presence of multiple formats I_l^{MF} would not be needed in the potential function $f(\cdot)$. The resulting specification for Π_{fkl}^z would be

$$\Pi_{fkl}^z = f(X; \alpha, \lambda, \delta) * h(N_k, \dots, N_{k'}, \dots, N_K; \beta, \tau) * \xi^z * (v_{fl})^{\rho-c} * \eta_{fl} \quad (13)$$

If stores cannot be distinguished by formats, all stores are homogenous, and the competitive effect $h(\cdot)$ is due only to the total number of rivals, and consumers’ preference for store formats (λ^k) and the format-specific scaling factor (τ_k) are not needed. The resulting specification for profits, Π_{fl}^z , would be

$$\Pi_{fl}^z = f(X; \alpha, \delta) * h(N; \beta) * \xi^z * (v_{fl})^{\rho-c} * \eta_{fl} \quad (14)$$

Thus, the simpler models are embedded in our model.

2.6 Equilibrium Choice Probabilities of Firms

²⁷ We continue to use the multiplicative form here for simplicity. Taking log transformation, as we do subsequently, gives the profit function the more familiar form of (profit = revenue – cost)

In our static set-up, all firms simultaneously make their entry and location decisions. When making a decision, each firm predicts the number of rivals of various formats that might enter a retail zone and choose a certain retail location. A firm doesn't observe its rivals' idiosyncratic component of cost (η_{jt}) but it knows the distribution, so it can form a conjecture or assign a probability to a rival's choice of a location. Summing over all such locations within a distance band, b , gives the firm's prediction of the total number of rivals of type t that will enter in distance band b of location l , $[N_{il}^b]$. Thus, if we drop the format specific parameter and price shocks, the expected competitive effect of different rivals in different distance bands is given by:

$$h(.) = \prod_b (1 + E[N_{kl}^b])^{\beta_{kb}} \prod_b \prod_{k' \neq k} (1 + E[N_{k'l}^b])^{\beta_{k'b}}, \quad (15)$$

where

$$E[N_{kl}^b] = N^z * \sum_{j \in L^{bm}} p_{kj}^z \quad (16)$$

is the predicted number of rivals of format k that choose a location in distance band b when N^z rivals enter the retail zone z , and p_{kj}^z is the probability of choice of location j (which lies in distance band b), by a firm of format k , in the retail zone.²⁸

Similarly, on the basis of its conjectures about rivals' and its own actions, the firm can predict the revenue potential of the location from consumers' shopping behavior. Hence, Equation (2) transforms into:

$$U_{il}^k = \lambda_i^k - T_{il} - \gamma E[h_{kl}] + \alpha_6 bus_l + \alpha_7 bus_l^2 + \alpha_8 E[N_l] + \alpha_9 E[I_l^{MF}] + \omega_{il}. \quad (17)$$

In equilibrium, when a firm makes its optimal choices, the conjectures of each firm about the strategies of all rival firms matches those rival firms' actual equilibrium strategies. This matching results in a nested fixed-point problem in the form of a mapping of the firm's strategy (which is a function of the competitors' strategies) onto competitors' strategies (which is a function of the firm's strategy).

To illustrate the fixed point problem, we drop the format specific parameter in the competitive effect and the unobserved demand and cost factors at a location, and then take the log transformation of the firms' profit function, Equation (12):

²⁸ One should be careful about how one counts the number of rivals. When considering the competition effect from locations in the same retail zone as the firm, the number of rivals is the number of firms that enter the retail zone, z , minus one. But when considering the competition effect from locations in a neighboring retail zone, z' , the number of rivals is exactly the number of firms that enter retail zone z' .

$$\ln(\Pi_{fkl}^z) = \delta^k \ln(GSP_{kl}) + \sum_b \beta_{kb} \ln(1 + N_{kl}^b) + \sum_b \sum_{k' \neq k} \beta_{k'b} \ln(1 + N_{k'l}^b) + \ln(\xi^z) + \ln(\eta_{fl}) \quad (18a)$$

For ease of exposition, we denote the left-hand side as $\tilde{\Pi}_{fkl}^z$ and group the potential and competitive effects together, denoted as $\tilde{\Pi}_{kl}$, such that

$$\tilde{\Pi}_{kl} = \delta^k \ln(GSP_{kl}) + \sum_b \beta_{kb} \ln(1 + N_{kl}^b) + \sum_b \sum_{k' \neq k} \beta_{k'b} \ln(1 + N_{k'l}^b). \quad (18b)$$

Also, by denoting the idiosyncratic component as ε_{fl} and the market-specific component as $\bar{\xi}^z$, we can rewrite Equation (18a) as

$$\tilde{\Pi}_{fkl}^z = \tilde{\Pi}_{kl} + \bar{\xi}^z + \varepsilon_{fl}. \quad (18c)$$

We assume that the idiosyncratic component follows a Type 1 extreme value distribution and is independent across firm-location combinations.²⁹ Moreover, we normalize the profitability of not entering the retail zone to 1 (i.e., log transformation equals 0).³⁰ Let the vector of the equilibrium location choice probabilities $[p_{k1}^z, p_{k2}^z, \dots, p_{kL^z}^z]$ of rivals of type k (k') be denoted by P_k^{z*} ($P_{k'}^{z*}$). Because the retail zone specific effect $\bar{\xi}^z$ is common to all locations, it does not influence the choice of locations after the firm has decided to enter the retail zone. The probability of the choice of a location l by a firm entering as type k thus is given by the logit form:

$$p_{kl}^z = \left(\frac{\exp(\tilde{\Pi}_{kl}^z)}{\sum_{k=1}^K \sum_{g=1}^{L^z} \exp(\tilde{\Pi}_{kg}^z)} \right) \forall l = 1, \dots, L^z, \forall k = 1, \dots, K. \quad (19)$$

This system of $(K * L^z)$ equations define the equilibrium location choice probabilities as a fixed point problem. As the probabilities of choosing a format and location, conditional on entry, must add up to 1, we essentially have a system of $(K * L^z - 1)$ equations in $(K * L^z - 1)$ unknowns. Since firm's strategies are continuous in its rivals' strategies, by Brouwer's fixed point theorem, at least one solution exists for this system of equations.

²⁹ This assumption does not allow for correlations in profitability across firms for the same location or across locations for the same firm. Such considerations are beyond the scope of this research.

³⁰ This normalization may be rationalized as a fixed-entry cost that is symmetric across all firms.

The total number of firms from the pool of F potential entrants that enter the market is determined by the attractiveness of the locations in the market for all firms and the unobserved retail zone effect $\bar{\xi}^z$. Because the profitability of not entering the market is normalized to 1 (log transformation equals 0), the probability that a firm enters the retail zone is given by:

$$\Pr^z(Entry) = \left(\frac{\exp(\bar{\xi}^z) * \sum_{k=1}^K \sum_{l=1}^{L^z} \exp(\tilde{\Pi}_{kl}^z)}{1 + \exp(\bar{\xi}^z) * \sum_{k=1}^K \sum_{l=1}^{L^z} \exp(\tilde{\Pi}_{kl}^z)} \right). \quad (20)$$

Hence, the total expected number of entrants in the retail zone, $(N^z + I)$, is given by:

$$(N^z + I) = F * \Pr^z(Entry). \quad (21)$$

By fixing the potential number of entrants F and observing the actual number of entrants $(N^z + I)$, $\bar{\xi}^z$ can be estimated as

$$\bar{\xi}^z = \ln(N^z + I) - \ln(F - N^z - I) - \ln\left(\sum_{k=1}^K \sum_{l=1}^{L^z} \exp(\tilde{\Pi}_{kl}^z)\right). \quad (22)$$

We assume that $\bar{\xi}^z$ follows a normal distribution, $N(\mu, \sigma)$, whose parameters can be estimated on the basis of the vector of $\bar{\xi}^z$ across the set of retail zones. Although the pool of potential entrants F is not observed, varying the size of this pool should have only a miniscule effect on the location choices, because $\bar{\xi}^z$ adjusts accordingly, relative to the outside option's fixed effect (i.e., choosing not to enter).³¹ Hence, we fix F to be a large number for all markets ($F = 40$).

Because the probability of entry is a function of the firms' equilibrium conjectures, which in turn is a function of the number of firms that enter the retail zone, a simultaneous solution for this set of equations provides the joint equilibrium predictions for the location choice probabilities and the number of entrants.

2.7 Estimation

We consider three store formats in our application and thus tailor our model to three formats ($K = 3$). We apply the Nested Pseudo Likelihood (NPL) approach to estimate the

³¹ See discussion in Seim (2006)

equilibrium location choice probability vectors, $P_1^{z^*}$, $P_2^{z^*}$, and $P_3^{z^*}$, of the firms in each market, and the model parameters.

In the NPL approach we start with an initial guess for the conditional choice probability vectors in each retail zone. For each retail zone, we observe the number of firms that enter and so we use the observed proportion of firms that choose a location as our initial guess. Starting with an initial guess for the parameter vector θ ($= [\lambda, \alpha, \delta, \beta, \tau, \rho, c, \mu, \sigma]$), we can predict consumers' shopping behavior, the prices, and the revenue at each retail location, based on the choice probabilities. We also estimate the prices and revenues of stores for the observed store configuration. Comparing our price estimates with the observed prices of the store chain for which we have price data, we get the price shocks, ζ_{jl} and hence the distribution of these shocks. Similarly, comparing our estimated revenues with the observed revenue for all stores at the locations that were chosen, we get the revenue shocks of firms, ν_{jl} , in these locations and hence the parameter ρ for the distribution of revenue shock.³² Now, for a draw of these shocks across all retail locations, we calculate the profit potential at each location. Then we calculate the elements of the probability vectors as given by Equation (19) and the vector $\bar{\xi}^z$ as in Equation (22). We do this for each Halton draw of the price shocks and revenue shocks, and then take the mean choice probabilities and retail zone specific scaling factors across a large number of Halton draws.

The probabilities of the observed location choices of firms, stacked across firms and retail zones, enters the conditional likelihood function:

$$L(\theta_l | \mu, \sigma, \rho, \zeta) = \prod_{z=1}^Z \prod_{k=1}^K \prod_{l=1}^L (p_{lk}^z(\theta_l, X, N^z))^{I(l=l_k)} \quad (23)$$

where, $\theta_l = [\lambda, \alpha, \delta, \beta, \tau, c]$, $I(l = l_k) = 1$ if location l is chosen by a firm of format k and is 0 otherwise. This is the likelihood, conditional on the unobserved revenue shocks at the chosen locations and the retail zone specific scaling factors.

The unconditional likelihood function is given by:

³² We acknowledge that since we only observe the revenue for firms at the locations that were chosen, there is a potential selection problem. Elickson and Misra (2007) show that this problem can be avoided by using a control function. We are currently exploring this solution.

$$L(\theta) = L(\theta_1 | \mu, \sigma, \rho, \zeta) \\ * \prod_{z=1}^Z \prod_{k=1}^K \prod_{l=1}^L \left(\phi \left(\left(\log(R_{fkl}^0) - \log(\hat{R}_{fkl}) \right), 0, \rho \right) \right)^{I(l=l_k^*)} \left(\phi \left(\left(\log(h_{fkl}^0) - \log(\hat{h}(\cdot)_{kl}) \right), 0, \zeta \right) \right)^{I(l=l_k^*)} * \phi(\bar{\xi}^z, \mu, \sigma) \quad (24)$$

where, $\phi(\cdot)$ is the pdf of the normal density, R_{fkl}^0 is the observed revenue of a store of format k at location l , \hat{R}_{fkl} is the predicted revenue that we obtained in Equation (9), h_{fkl}^0 is the observed price and $\hat{h}(\cdot)_{kl}$ is the estimated price.

2.8 Identification

It is useful to relate our identification mechanism with the existing entry-location models where the inferences about firms' trade-offs are made through the observed entry and location decisions of firms. Seim (2006) used two kinds of variations in the data to identify the effects of demand characteristics (such as, population and per capita income) and competition – variation in demand characteristics, conditional on number of rivals around a location and variation in number of rivals around a location, conditional on demand characteristics.

To the extent that the volume and price effects of competitors affect store revenues, we have decomposed the reduced form profit function into its components of revenue and cost to isolate the effects of competitors. Further, we have decomposed the revenue into its components of volume and price. However, since the location of competitors affects both the volume and the price, complete non-parametric identification of the two effects would ideally require data on sales or prices at all stores. Currently, we only have price information for a set of stores belonging to one store chain. This is interesting in a way because it is a more realistic situation as firms usually have more information about themselves than about others. While using this piece of information aids non-parametric identification of the competitive effects of rivals, we partly rely on our functional form assumption on how the locations of rivals affect store volume versus how they affect store prices.³³ Specifically, for the price effect, we follow a similar approach as in Seim (2006) and allow prices to be affected by the number of rivals and their distances from a location. For the volume effect, we model consumers' shopping location choice behavior, which incorporates the number of firms and their geographic location around consumers. Hence, in

³³ We are in the process of trying to obtain additional price data for other stores.

addition to the two kinds of variations in the data mentioned above, we exploit another kind of variations in the data - variation in the number and locations of stores around consumers.

3 Data and Spatial Aspects of the Model

Here, we discuss the data that we use for our analysis and the spatial aspects of the model.

3.1 Data and Markets

We have data on grocery store locations (latitude and longitude), weekly store revenues and store formats for the period 2007-2008, which were obtained from ‘*Trade Dimensions*’.³⁴ We classify grocery stores into three types; Supermarket format (SMF), Superstore format (SSF) which includes Superstores and Super Centers, and Other store format (OSF) which includes Limited assortment stores, Wholesale clubs, Natural foods / Gourmet stores and Food / Drug combination stores. We use demographics data at the CBG level, obtained from the Census. The US Census has decennial census data for the CBG level (for the years when a national census was held) but annual census projections only at the county level. The last national census was held in 2000. Hence, we scale the 2000 census data, at the CBG level, by the proportion of change at the county level to get estimates of demographics for 2007.

We introduce a new approach to obtain zoning data across markets on a national scale, from a digital land cover data called *National Land Cover Dataset* (NLCD). Like the US census data, the latest digital maps of NLCD correspond to the year 2001. However, spatial zoning regulations, by their very nature, are relatively stagnant over a period of 7 years and rezoning is usually rare and minuscule. Hence, we feel comfortable applying the same data for the year 2007.³⁵ In the next sub-section, we describe how we collect the zoning data and how we define retail zones and retail locations.

We have store data for the entire nation and, by using our approach, zoning data also can be collected for any number of markets. However, an analysis of store location choice at a national scale makes the state space of retail locations extremely large. Thus, for our application, we focus on three north eastern US states (New York, Pennsylvania and Ohio) to keep the size of

³⁴ We thank Nielsen for providing us with this data.

³⁵ For the markets that we analyze, we also have data on store locations (but not store revenue) for the year 2001. Between these two data sets we see that the number of stores have increased by 12% but less than 0.5% of the stores (2 out of 440 stores) were located outside any of our retail zones. These stores were found to be relatively small Food and drug stores and were excluded in our analysis.

the problem manageable. Within these US states we analyze firms' entry and location choice decisions in retail zones of small and medium sized towns. For large cities, like, say, New York city, retail zoning regulations tend to be very complex and fairly intertwined with residential zoning. Sub-division zoning makes it is difficult to identify distinct retail zones in these markets.³⁶ Also, in these markets multiple stores of the same store chain may exist in close proximity, in which case one would have to worry about cannibalization. Hence, in our application, we focus only on small and medium sized towns. This leaves us with a total of 107 towns, accounting for 438 grocery stores.

Finally, we also have price data for a set of stores that belong to one store chain. This is a large enough store chain such that it is present in 75% of the towns and accounts for 20% of the stores in these markets. We use sales weighted price index as a proxy for a store's pricing ability.

3.2 Spatial aspects of the model and zoning data

3.2.1 Consumer locations and Retail locations

As discussed in the model section, an analysis of store location choice requires us to discretize the continuous geographical space. A natural choice for a discrete unit of geographical area is a census division (e.g., CBGs or Census Tracts). Previous models of firms' location choice assumed that all consumers and stores are located at the population weighted center of their respective census divisions. There are two drawbacks to this assumption with regards to store locations: (1) Stores are rarely present on the interior of a census division, rather, they are present on roads that border these census divisions (See **Figure 3a**); (2) census divisions vary extensively in size so that, for large census divisions, stores may be located quite far from the center (See **Figure 3b**). Thus, if stores are assumed to be at the center of census divisions, it could seriously distort the distances between competing stores and the distances consumers have to travel.

To address these issues, we improve upon the traditional approach and distinguish between consumer locations and retail locations. Like previous research, we assume that consumers are located at the population weighted center of CBGs because we do not have detailed information about consumers beyond the CBG level. These are our *consumer locations*.

³⁶ One alternative is to collect separate zoning data for individual large cities from their respective municipalities and augment our zoning data.

However, to ensure minimal distortion of store locations, we discretize the geographical space into units of 1 sq. mile block units and assume that firms make their location choice decisions among those units where zoning permits them to set up a store. That is, we divide towns into a grid of 1 sq mile blocks and then identify which of the 1 sq mile blocks in a market are ‘potential’ *retail locations* based on zoning restrictions. Not every 1 sq. mile block land area is a potential retail location because depending on the zoning in the area, a 1 sq mile block land area could be a residential or industrial zone. We group geographically contiguous ‘potential’ *retail locations* into *retail zones*.

3.2.2 Zoning data

Digital land-cover data are among the most popular data sources used in ecological research and natural resource management but so far they have not been exploited in market structure research. For our application, we identify retail locations by using a new digital land-cover dataset called National Land Cover Dataset (NLCD) in conjunction with GIS tools such as *ArcGIS* and *Google Earth*.

Multi-Resolution Land Characteristics (MRLC) consortium, a conglomerate of several Federal agencies, has produced *NLCD* data in 1992 and 2001, referred to as *NLCD 1992* and *NLCD 2001*. These provide a relatively consistent and accurate digital land cover data for the coterminous United States. *NLCD 1992* was the first national land-cover mapping project that applied a 21-class, geo-referenced, land cover classification. It is derived from the early to mid-1990s *Landsat Thematic Mapper (TM)* satellite data (see Vogelmann et al., 2001). *NLCD 2001* is the second project which updated the data for the year 2001 (see Homer et al., 2004).³⁷ Both these datasets have a spatial resolution of 30 meters. That is, every 30 sq. meter area of land is classified as a specific land type and is allocated one pixel point with a specific color code and with the corresponding latitude-longitude.³⁸

To identify retail locations and retail zones, and the retail business density in retail locations, we proceed in the following steps.

³⁷ *NLCD 1992* has classifications for low intensity residential and high intensity residential zones. It combines commercial and industrial zones into one classification. *NLCD 2001* has classifications for low intensity residential and medium intensity residential zones. It combines high intensity residential, commercial and industrial zones into one classification.

³⁸ A pixel point is one of the individual dots that make up a graphical image. Each pixel point combines red, green, and blue phosphors to create a specific color.

Step 1: Markets

In the first step, we use *NLCD 2001* to extract data on residential, commercial and industrial land areas corresponding to the towns in our data. That is, we extract the pixel points corresponding to these land classifications (See **Figure 4a**). This gives us the geographical space we are interested in. Now, we need to discretize this geographical space into units of 1 sq. mile blocks. One way to do this is to divide each town into a grid of 1 sq. mile blocks. This however is hard to do because towns do not have any uniform shape. Hence, we first enclose the towns into distinct rectangular shapes which can be easily divided into a grid. That is, we construct imaginary rectangular borders (north, east, south and west borders) for the towns and call each such rectangular area a ‘*market*’. This enclosure of towns into rectangular markets is done based on our visual inspection of the pixel density of residential, commercial and industrial land areas (See **Figure 4b**). When it is difficult to distinguish between two neighboring towns (i.e., it is hard to construct a border between them from simple visual inspection) because say, their residential or commercial land areas heavily overlap, we group them together and enclose them in a larger rectangular area.

The construction of these rectangular markets is only to aid the discretization of the geographical space. The choice of this shape is innocuous and it has little impact on our model or estimation since we allow consumers to shop across markets (up to a radius R) as well for firms to compete with rivals in neighboring markets (again, up to a radius R).

Step 2: Retail locations

Having obtained a rectangular market, we then divide it into a grid of 1 sq mile blocks (See **Figure 4c**). Next, we identify which of these square blocks are ‘potential’ retail locations.

Since the *NLCD* datasets were developed for ecological research and natural resource management, they were more concerned about detailed classifications of vegetation and barren land types. *NLCD 1992* separates high density residential land from commercial and industrial land whereas *NLCD 2001* combines them. Hence, we match the two data sets using *ArcGIS* software to extract the commercial and industrial land data for 2001 (See **Figure 4d**). We believe that this is an innocuous step as it is unlikely that a region that was a commercial or industrial zone in 1992 would convert into a high density residential zone by 2001.

Next, to identify ‘potential’ retail locations, we project this extracted data, and the grids of 1 sq. mile blocks, on to *Google Earth*. *Google Earth* has a tool which allows one to find

‘Places of Interest’ in any region of the US. These *‘Places of Interest’* include banks and ATMs, bars and clubs, religious places, major grocery stores, etc (See **Figure 4e**). Hence one can check for the presence of any kind of businesses at a location. We carefully comb through our grid of 1 square mile blocks and, among the commercial and industrial land areas, we check for the presence of any major retail stores, grocery stores and shopping centers (See close-up in **Figure 4f**). All 1 sq. mile blocks with any major retail stores or grocery stores or a shopping center are then defined as ‘potential’ retail locations.

Note that not all square block units with commercial and industrial land pixel points would be retail locations. Some of them may be purely industrial land areas and some others may be commercial land areas but only with small businesses like banks, restaurants, etc. Thus, even within commercial zones, it is important to distinguish between a typical commercial land area and land areas available for entry by major retail stores.

Step 3: Retail zones:

Finally, we define a retail zone as a collection of retail locations. We form distinct clusters of retail locations such that the minimum distance between any two retail locations belonging to separate clusters or retail zones is more than 2 miles.

Note that there could be multiple retail zones within a single market (See **Figure 4g**). As discussed in the model section, certain features of individual retail zones (major intersections, parking lots) might make it more attractive so that firms choose one retail zone over another within the same market. When the researcher does not observe such retail zone specific features, a homogeneous market specific effect would not control for endogenous entry into a retail zone. Hence, our zoning data allows us to improve on the traditional approach of modeling firms’ entry decisions into an entire market and instead model entry decisions into retail zones.

Step 4: Business density and business density weighted center of retail locations

From the NLCD data, we use the number of pixel points in each 1 sq mile retail location block as a measure of the retail business density in that location. The mean of the latitudes and longitudes of the pixel points within a retail location gives us its business density weighted center.³⁹ This is a reasonable approximation because the pixel data is fairly accurate and there is

³⁹ Each pixel represents a 30 meter x 30 meter (900 sq meter) area, and a 1 sq. mile area will have a maximum of around 2800 pixel points.

negligible overlap between these pixel points and what may be residential land within the same 1 sq. mile block area (See **Figure 4h**).

4. Results

In this section we discuss the results of our estimation. Table 3 presents the estimates for the consumer or the demand side of the model. As expected, consumers experience a negative travel cost that is convex with respect to distance (The coefficient of d_i^2 is negative and significant at the 1% level). This travel cost is lower for consumers who are far away from the nearest retail location. Hence, consumers who are far away from the nearest retail location are more willing to travel long distances to get to a grocery store.

Demographic characteristics seem to have very little explanatory power for consumers' preferences for individual store formats. But consumers in rich neighborhoods (i.e., high median household income) prefer the Supermarket format or the Other store format more than the Superstore format. Conversely, consumers who are far from the nearest retail location and have to travel a lot prefer the Superstore format (which allows them to combine purchases across a larger set of categories) more. These findings are quite intuitive.

Our estimates show that consumers not only value economies of scope from the presence of other, non-competing, businesses at a location but they also value the agglomeration of multiple grocery stores at the location. Both store agglomeration and format agglomeration parameters are positive and significant at the 1% level. Hence, consumers value store agglomeration even more when the cluster of stores have different formats. Consequently, store and format agglomeration increase consumers' propensity to shop at a location, thus increasing volume at that location. Finally, the income parameters indicate that firm revenue increases with population and income but that this relationship is concave.

Next, we present the results of the supply (firm) side of the model in Table 4. We considered competition from rival stores in three distance bands – 0-2 mi. and 2-4 mi. and 4-6 mi. None of the estimates for the 4-6 mi. distance band were significant.⁴⁰ Hence, we only present the results for the first two distance bands. We find that, as expected, stores with Superstore format compete a lot more than the other two formats. Interestingly, intra-format competition is almost twice as much as inter-format competition. Among the different inter-

⁴⁰ This is expected because stores are unlikely to compete with each other when they are more than 4 mi. apart.

format competition effects, supermarkets and superstores compete the most with each other but this is still far lower than the intra-format competition for these two formats. This suggests the importance of accounting for format differentiation, in addition to spatial differentiation.

Finally, the logarithm of price shocks follows a normal distribution with a standard deviation of 1.8423 and the logarithm of revenue shocks across locations follows a normal distribution with a standard deviation of 1.1213. The cost shocks are highly correlated with the demand shocks, with a significantly high correlation of 0.9546. Thus, even though a location may have high revenue potential, we may not find multiple stores at that location due to the associated high costs. Only when the location is extremely attractive (i.e., it gets an extremely high price shock or revenue shock), we may observe store co-location that is not strategic.

5. Counterfactual Simulations

We report two counterfactual simulations which help assess the relative importance of zoning and agglomeration effects. Figure 6a shows the distribution of inter-store distance in the actual data across the 107 towns that we used for estimation. We see that 60% of stores locate within 1 mile of each other. When we turn off agglomeration and zoning constraints (figure 5b), only 28% of stores co-locate within 1 mile of each other. This level of co-location may be due to concentration of high demand or low cost. With only agglomeration turned on (figure 5c), 43% of stores located within 1 mile of each other, suggesting that agglomeration effects have a significant impact on observed co-location. In contrast, when only zoning is on, we find only 31% of stores are located within 1 mile apart (figure 5d), suggesting on first glance that zoning has very limited direct impact on observed co-location.

What is interesting, however, is that while zoning has limited direct impact, the interaction between zoning and agglomeration effect is extremely high because co-location increases to 60% when both effects coexist (figure 5a). Thus the impact of zoning on co-location is high only when agglomeration benefits are high.

Why do we see such interaction effects between zoning and agglomeration benefits? To understand these effects, we perform counterfactual analysis in two markets: one with less restrictive and another with more restrictive zoning. The optimal locations are shown in Figure 6. In the less restrictive zoning setting, we find that stores are located at the extremes of the zone, suggesting that zoning restrictions constrain the extent of spatial differentiation in this market.

When we make zoning more stringent, one would therefore expect that stores would continue to be at the edges of the retail zones. However, the optimal locations reveal a surprising pattern. When zoning is restricted, we find that some stores actually agglomerate and the inter-store distances are lower. In retrospect, we can understand the logic of why this happens. When zoning is relaxed, stores can be more spread out allowing for benefits of differentiation to be large enough. However when zoning is restricted, firms cannot differentiate enough; this leads to a discontinuity where stores now recognize that by co-locating they can gain from agglomeration benefits which may outweigh the relatively constrained benefits from differentiation because of the tight zoning regulations. Thus zoning has discontinuous impact on co-location in the presence of agglomeration benefits. This explains the high interaction effect of zoning and agglomeration that we find based on the results in Figures 5a – 5d.

6. Conclusion

Spatial location choice is a problem of great interest to many firms. This paper develops a structural model of entry and location choice that helps disentangle the agglomeration-differentiation tradeoff. The paper was motivated by the high levels of spatial clustering observed in grocery markets. Specifically, we investigated the factors driving co-location. Is this due to the beneficial effect of co-location (competitor is a friend)? Or is it despite the margin suppressing effect of competition (competitor is an enemy) due to high demand or low cost at the particular location? Or could it be due to zoning?

The paper introduced three key innovations. To separate the differentiation-agglomeration effect, we use a modeling innovation by first decomposing profits into revenues and costs and then further decomposing revenues into its components of consumer choice based volume and competition based prices. To separate demand and cost explanations, we augment entry and location data with store revenue data. To non-parametrically identify the volume and price effects of competitors we include price data for one store chain. To capture zoning effects, we introduce a new approach to obtain zoning data that takes advantage of publicly available database (NLCD) and tools (ArcGIS and Google Earth). This paper is one of the early applications of the NPL approach in an empirical setting, especially within a static framework. This paper thus combines modeling and methodological innovations coupled with new data to improve upon extant models of location choice.

The consumer and competition model provided intuitively interesting results. Further, the results of our counterfactual analysis highlighted the value of developing a structural model. First we find that agglomeration effects are a key driver of observed location decisions, suggesting that counterfactual analysis based on omission of these effects in extant research may be suspect. Second, we find that zoning has little direct effect on co-location. However, zoning interacts with agglomeration to drive observe co-location. Such a nuanced understanding of how zoning regulations can affect spatial configurations can be possible only with structural estimation of the type we pursue in the paper.

The structural model we developed has important managerial applications. A retailer seeking to open a new store needs profit estimates for each location in a market to decide whether to enter the market and which location to choose. But the model has value even for existing stores. Retailers need estimates of expected profit given the location's demand and cost characteristics and competitive environment so that the firm can benchmark the store's actual performance against what should be expected. This can aid managers in making decisions about incentivizing store managers, decision on store closures and targeted marketing strategies. Further, the nuanced effects of zoning restrictions on spatial configuration because of agglomeration effects are clearly important for both managers and local governments.

We conclude with a discussion of limitations which offer suggestions for future work. First, our identification of the volume and price effects is partially aided by functional form assumptions on how locations of competitors affect volumes and prices differently. This is because currently, we only have price information for a set of stores belonging to one store chain. None the less, this is interesting because it is more realistic as firms usually have more information about themselves than about others. Secondly, we treat entry decision in a static equilibrium framework, even though a dynamic model may be more appropriate given that these decisions are made over time. Such a modeling approach requires better data (timing of entry and exits) as well as richer modeling framework to solve the dynamic game. Finally, we have treated store entry decisions across markets as independent, unlike recent work by Jia (2008), who models the chain entry decision. However her modeling approach is restricted to a small number of competing chains and is hard to extent to our supermarket setting involving a larger number of competitors. The chain issue is important and awaits future research.

References

- Arentze, T.A., O. H. Oppewal, and H.J.P. Timmermans (2005), "A Multipurpose Shopping Trip model to Assess Retail Agglomeration Effects," *Journal of Marketing Research*, 42 (February), 109-115.
- Aguiregabiria, V., and Mira, P. (2007), "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1, 1 - 53.
- Aguiregabiria, V., P. Bajari, M. Draganska, L. Einav, D. Horsky, S. Misra, S. Narayanan, Y. Orhun, P. Reiss, K. Seim, V. Singh, R. Thomadsen, and T. Zhu (2008), "Discrete Choice Models with Strategic Interactions," Forthcoming in *Marketing Letters*.
- Bajari, P. and J. Fox (2006). "Complementarities and Collusion in an FCC Spectrum Auction," Working Paper
- Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2006), "Estimating Static Models of Strategic Interactions," *National Bureau of Economic Research*, Working Paper Series.
- Bajari, P., Benkard, L., and Levin, J. (2007a), "Estimating Dynamic Models of Imperfect Competition," *Econometrica* 75, 5, 1331 - 1370.
- Bajari, P., Hong, H. and Ryan, S. (2007b), "Identification and Estimation of a Discrete Game of Complete Information," *Technical report*, University of Minnesota.
- Bester, H. (1998), "Quality Uncertainty Mitigates Product Differentiation," *RAND Journal of Economics*, 29 (Winter), 828-844.
- Bresnahan, T., and P. Reiss (1991), "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 99, 977-1009.
- Chan, T. Y., V. Padmanabhan, and P. B. Seetharaman (2006), "An Econometric Model of Location and Pricing in the Gasoline Market," *Journal of Marketing Research*, Forthcoming.
- Ciliberto, F. and E. Tamer (2006), "Market Structure and Multiple Equilibria in Airline Markets," Working Paper.
- Datta, S., K. Sudhir, and D. Talukdar (2008), "The Differentiation-Agglomeration Trade-off in Location Choice," Working Paper.
- Draganska, M., Mazzeo, M., and Seim, K. (2007), "Beyond Plain Vanilla: Modeling Product Assortment Choices in the Ice Cream Market," Working Paper, Stanford University.

- Duan, A.J. and C. F. Mela (2007), "The Role of Spatial Demand on Outlet Location and Pricing," Working paper.
- Dudley, M. (1990), "Competition by Choice: The Effect of Consumer Search on Firm Location Decisions," *The American Economic Review*, 80 (5), 1092-1104.
- Ellickson, P.B., and S. Misra (2007), "Enriching Interactions: Incorporating Revenue and Cost Data into Static Discrete Games," Working paper.
- Ellickson, P.B., S. Houghton, and C. Timmins (2008), "Estimating Network Economies in Retail Chains: A Revealed Preference Approach," Working Paper.
- Fischer, J.H., and J.E. Harington (1996), "Product Variety and Firm Agglomeration," *RAND Journal of Economics*, 27, 281-309.
- Fox, J. (2007), "Semiparametric Estimation of Multinomial Discrete Choice Models Using a Subset of Choices," *The RAND Journal of Economics*, 38, 4, 1002 - 1019.
- Fox, E. J., S. Postrel and A. McLaughlin (2007), "The Impact of Retail Location on Retailer Revenues: An Empirical Investigation," Working paper
- Holmes, T. (2008), "The Diffusion of Wal-Mart and Economies of Density," Working Paper.
- Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan (2004), "Development of a 2001 National Land-Cover Database for the United States," *Photogrammetric Engineering & Remote Sensing*, 70, 7, 829 – 840.
- Jia, P. (2008), "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," forthcoming, *Econometrica*.
- Konishi, H. (2005), "Concentration of Competing Retail Stores," *Journal of Urban Economics*, 58, 488-512.
- Marshall, A. (1920), *The Principles of Economics* 8th ed., MacMillian, London.
- Mazzeo, M. (2002), "Product Choice and Oligopoly Market Structure," *RAND Journal of Economics*, 33, 221-242.
- Merrill Lynch, company reports, (2007).
- Pakes, A., M. Ostrovsky, and S. Berry (2004), "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)," *RAND Journal of Economics*, 38, 2, 373 - 399
- Pesendorfer, M., and Schmidt-Dengler, P. (2003), "Identification and Estimation of Dynamic Games," *National Bureau of Economic Research*, Working Paper Series.

- Ridley, D.B., F. A. Sloan and Y. Song (2007), "Retail Zoning and Competition," Working Paper.
- Seim, K. (2006), "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *RAND Journal of Economics*, 37 (3), 619-640.
- Shlay, A. B. and P. H. Rossi (1982), "Keeping up the Neighborhood: Estimating Net Effects of Zoning," *American Sociological Review*, 46, 703-719.
- Stahl, K. (1982), "Differentiated Products, Consumer Search, and Locational Oligopoly," *The Journal of Industrial Economics*, 31 (1-2), 97-113.
- Su, C., and K. L. Judd (2007), "Constrained Optimization Approaches to Estimation of Structural Models," Working paper
- Suzuki, J. (2008), "Land Use regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry," Working Paper.
- Thomadsen, R. (2005), "The Effect of Ownership Structure on Prices in Geographically Differentiated Industries," *RAND Journal of Economics*, 36, 4, 908 - 929.
- Thomadsen, R. (2007), "Product Positioning and Competition: The Role of Location in the Fast Food Industry," *Marketing Science*, Forthcoming.
- Varian, R. H. (1980), "A Model of Sales," *American Economic Review*, 70, 651-659.
- Vitorino, M. A. (2008), "Empirical Entry Games with Complementarities: An Application to the Shopping Center Industry," Working paper.
- Vogelmann, J.E., S.M. Howard, L. Yang, C.R. Larson, B.K. Wylie, and J.N. Van Driel (2001), "Completion of the 1990's National Land Cover Data Set for the Conterminous United States," *Photogrammetric Engineering & Remote Sensing*, 67, 6, 650 - 662.
- Watson, R. (2005), "Entry and Location choice in Eyewear Retailing," Working Paper.
- Wernerfelt, B. (1994), "Selling Formats for Search Goods," *Marketing Science*, 13 (3), 298-309.
- Wolinsky, A. (1983), "Retail Trade Concentration Due to Consumers' Imperfect Information," *The Bell Journal of Economics*, 14 (1), 275-282.
- Zhu, T. and V. Singh (2007), "Spatial Competition with Endogenous Location Choices: An Application to Discount Retailing," Working paper.

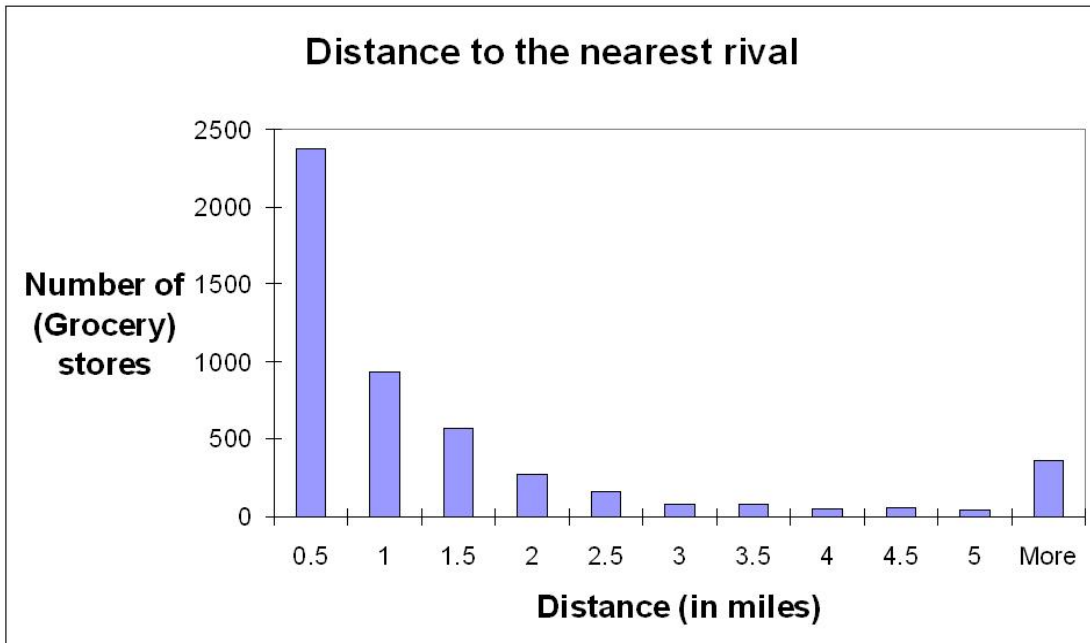


Figure 1: Over 47% of grocery stores are within 0.5 mi. of a rival store
(Data for 3 US states)

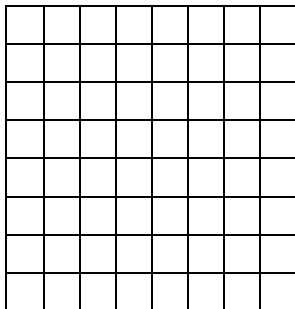


Figure 2a: A illustrative square market with the geographical space discretized into square blocks or 'locations'

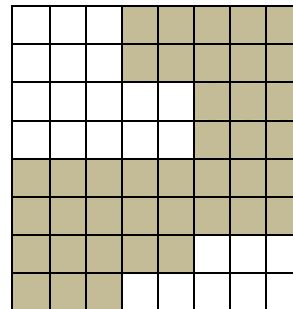


Figure 2b: Due to zoning regulations, firms can only choose among the 'potential retail location'
(White regions are 'Retail Zones')

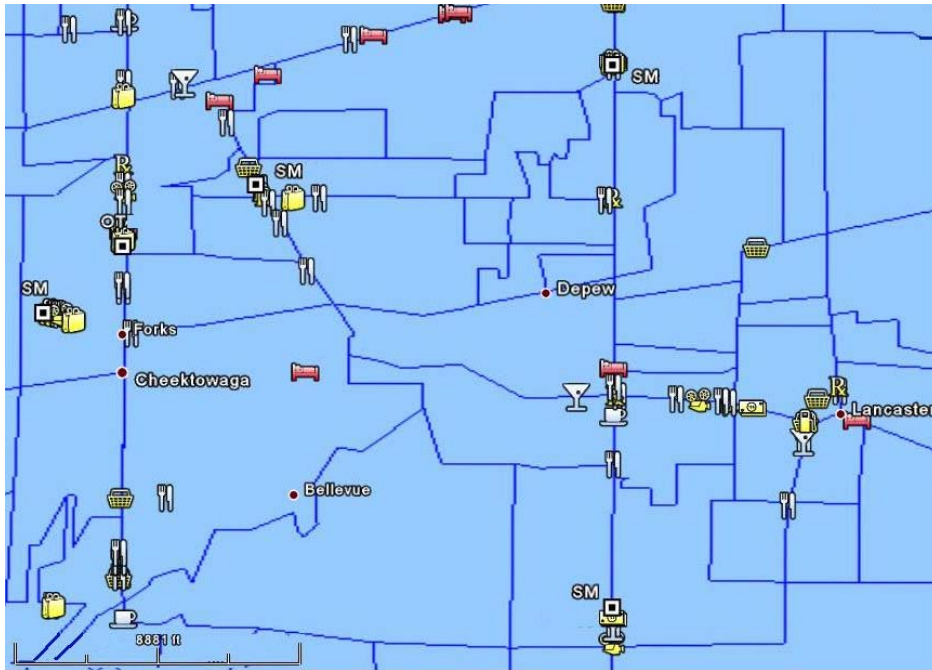


Figure 3a: Businesses are usually located on roads that border CBGs
 (Depew, NY; SM = Super Market Format)
 (Figure obtained by projecting CBG shapefiles from the Census on to *Google Earth*)

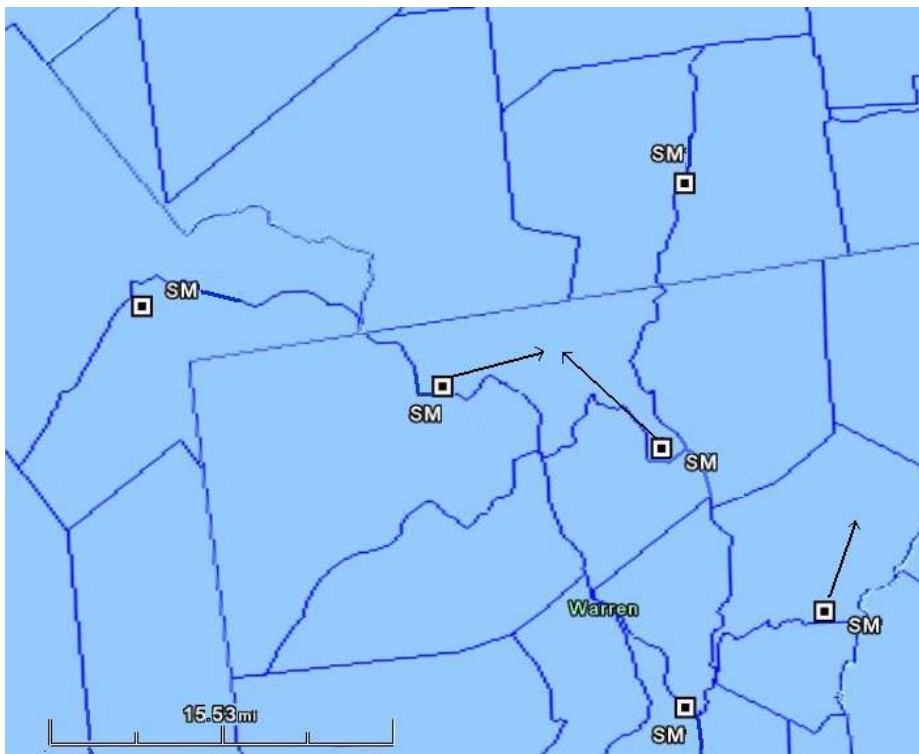


Figure 3b: In case of large CBGs, stores are far from center
 (Warren NY; SM = Super Market Format)
 (Figure obtained by projecting CBG shapefiles from the Census on to *Google Earth*)

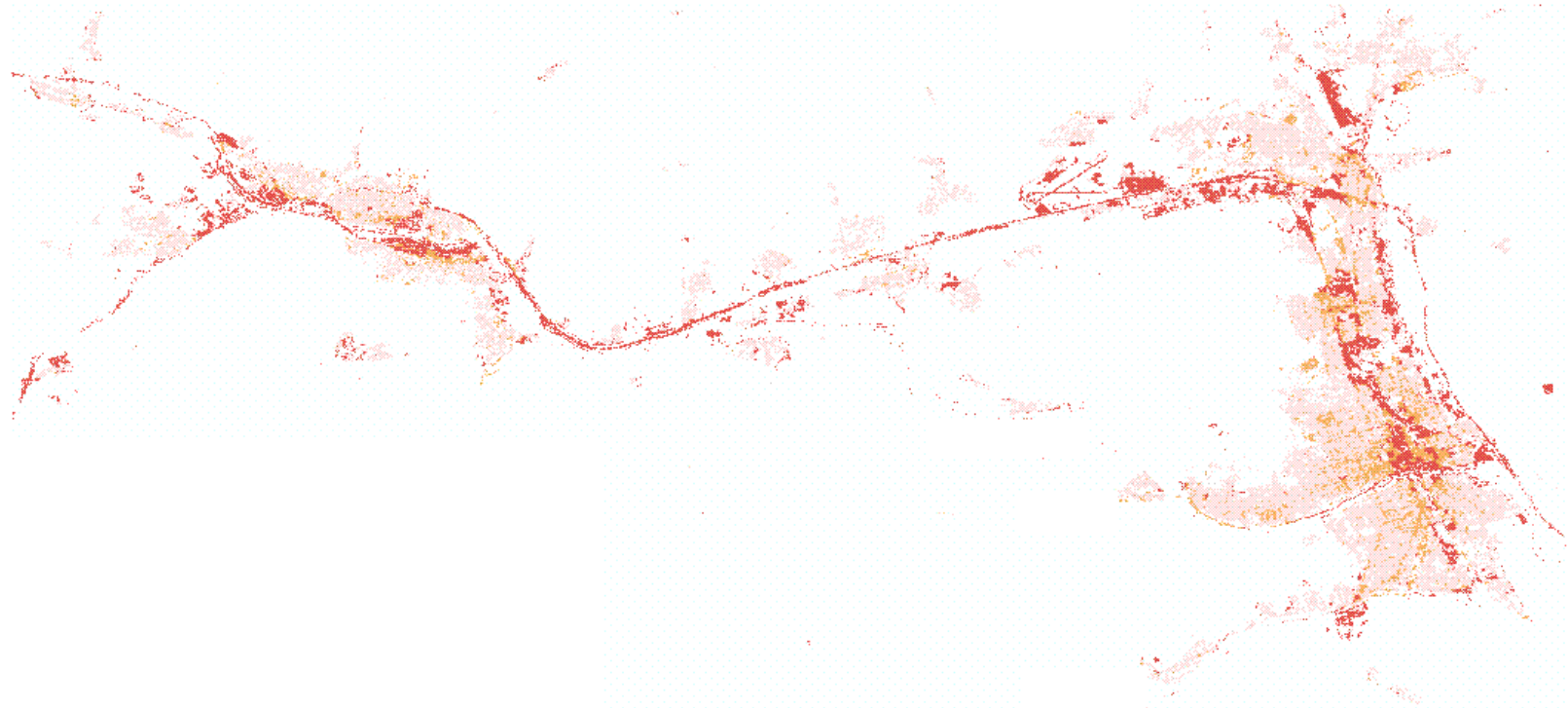


Figure 4a: Step 1; NLCD residential, commercial and industrial areas in three towns
(L to R): Corning, NY; Big Flats, NY; and Elmira, NY.

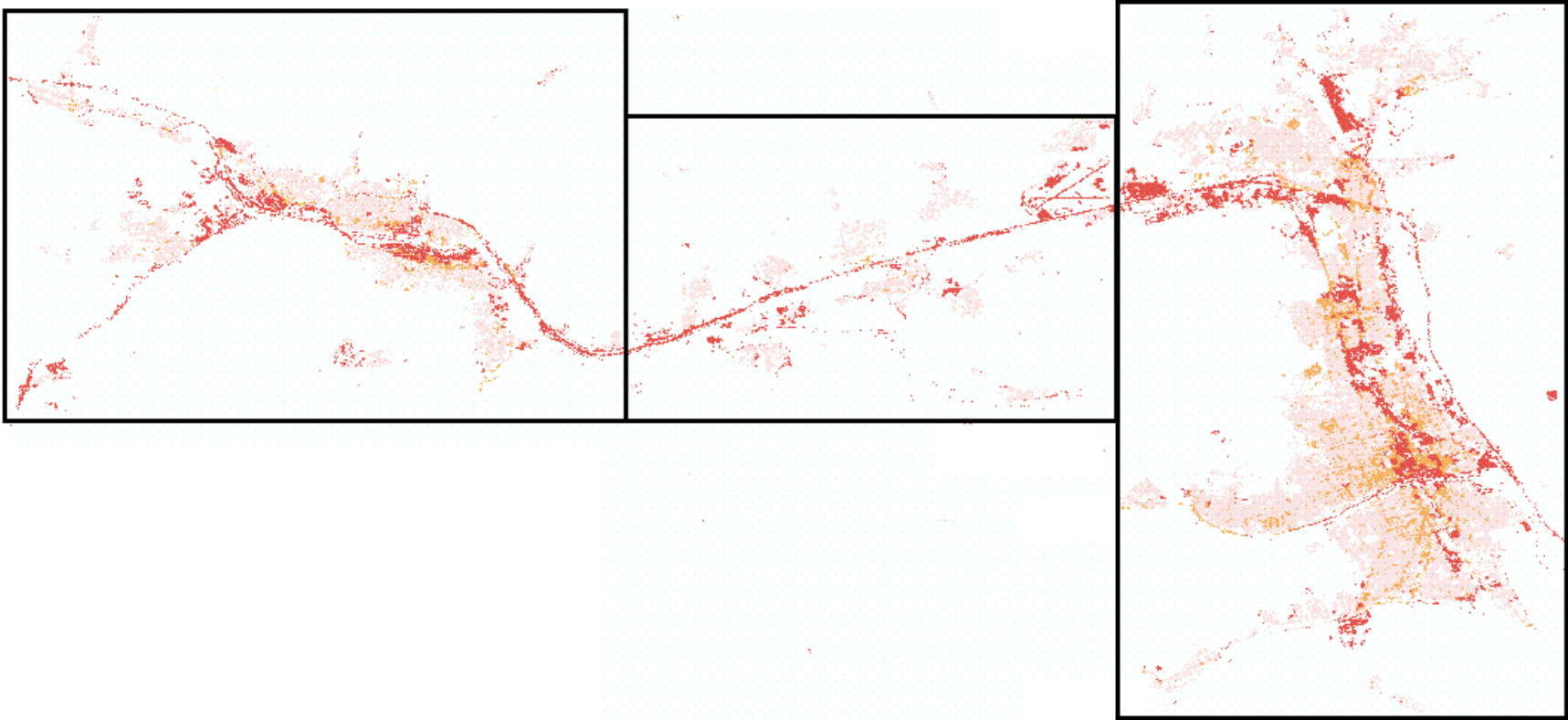


Figure 4b: Step 1; Defining rectangular markets (i.e., constructing imaginary north, east, west and south borders of towns) based on visual inspection of pixel density

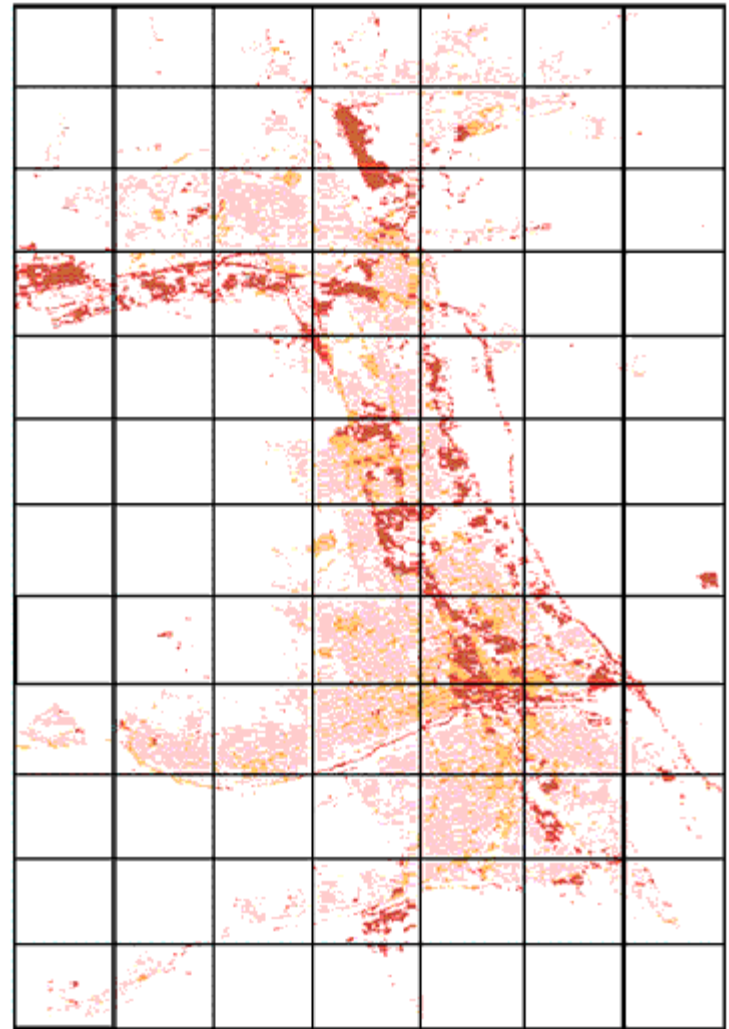
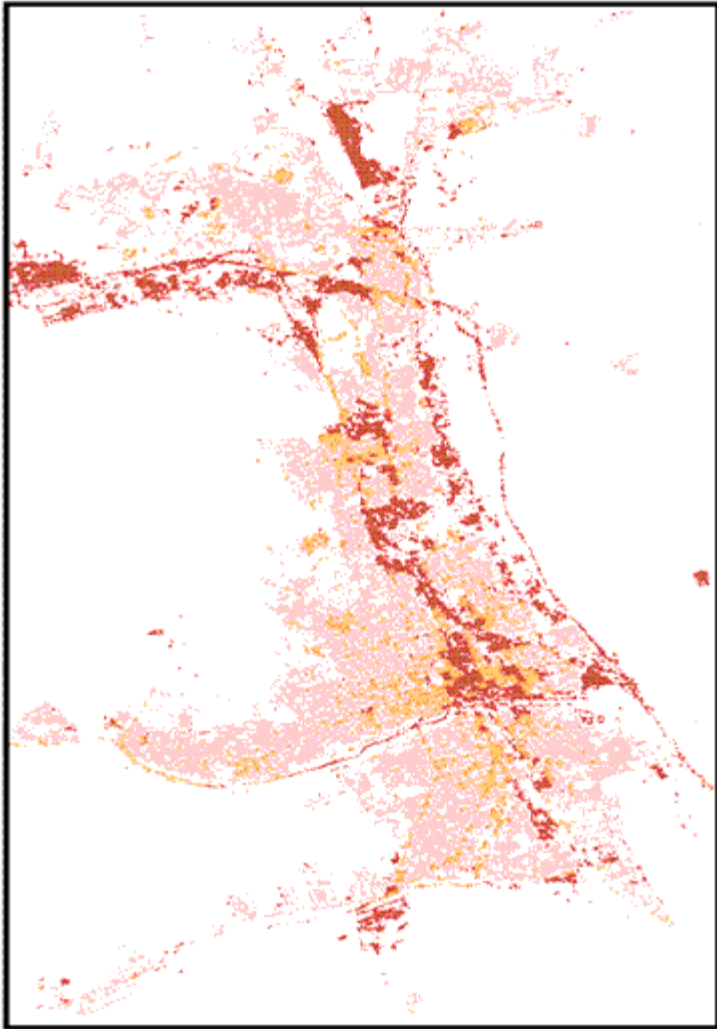


Figure 4c: Step 2; Discretizing the geographical space; Dividing a rectangular market into a grid of 1 sq. mile blocks (Elmira, NY)

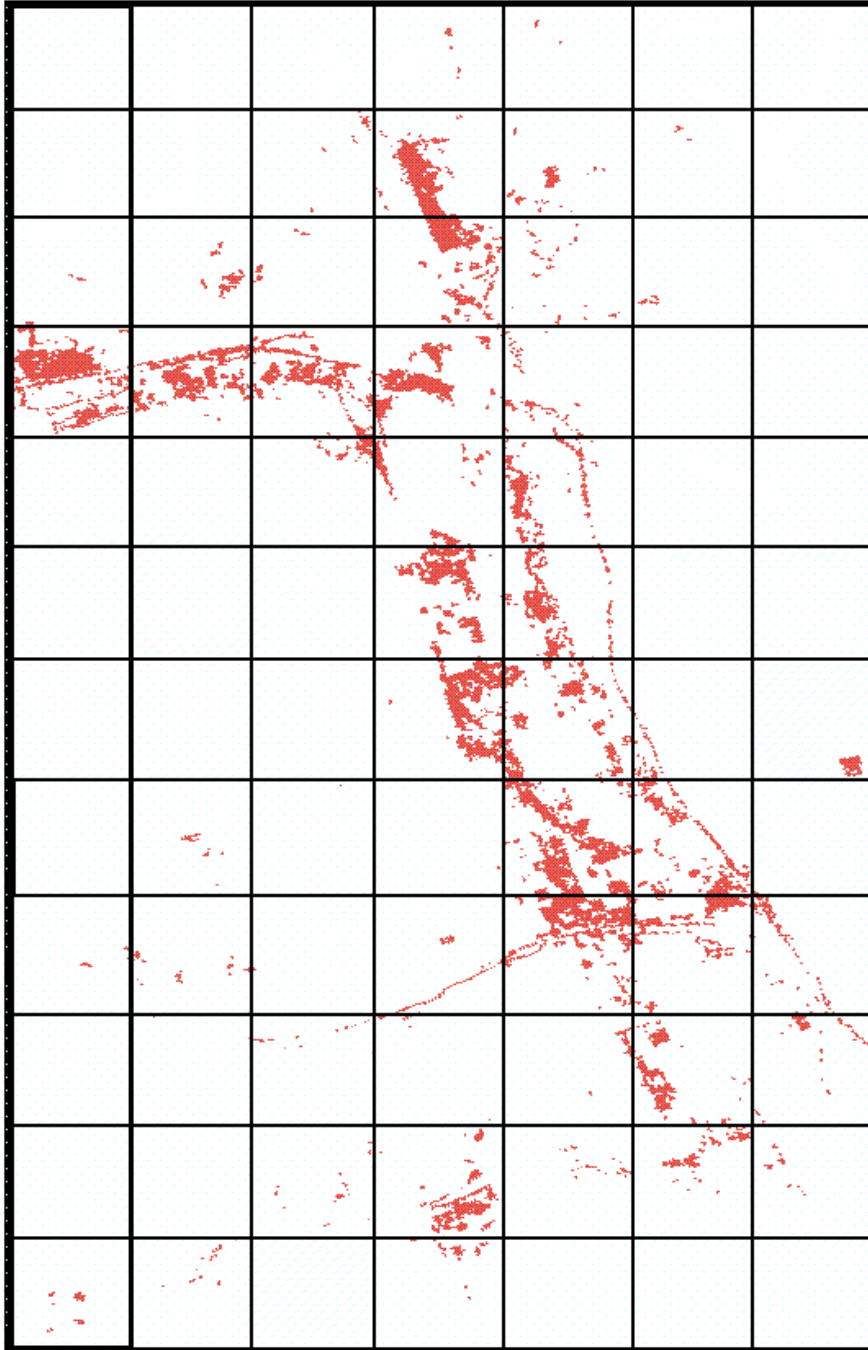


Figure 4d: Step 2; Commercial and Industrial land areas (Elmira, NY)

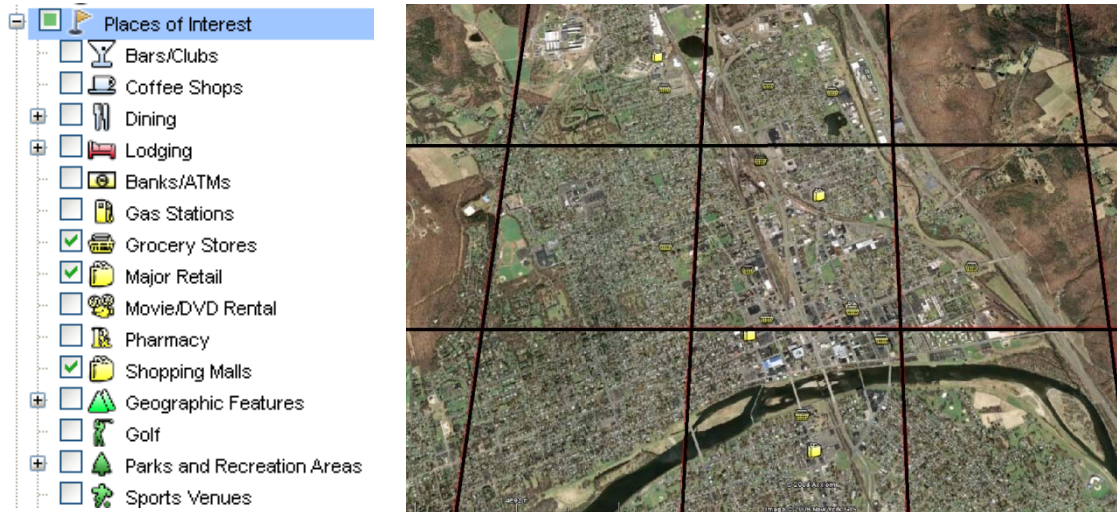


Figure 4e: ‘Places of Interest’ in Google Earth

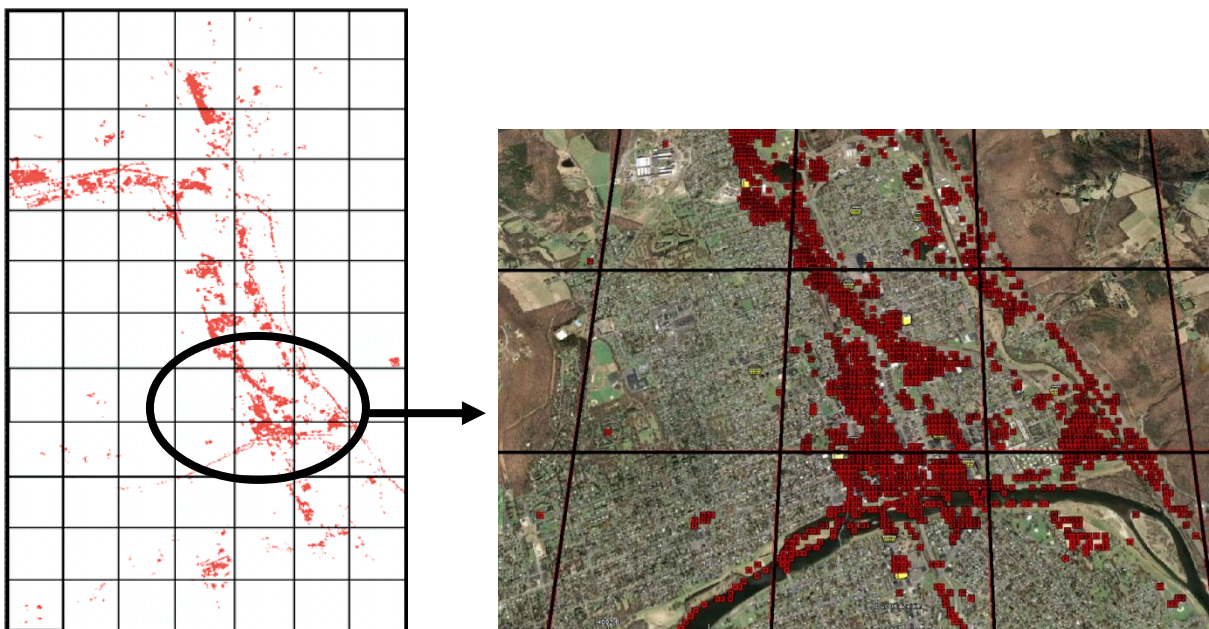


Figure 4f: Step 2; Projecting NLCD commercial and industrial land pixel points on to Google Earth. We carefully comb through the 1 sq. mile blocks corresponding to commercial and industrial land area to check for presence of grocery stores, major retail businesses and shopping malls in these blocks to identify ‘potential’ retail locations.

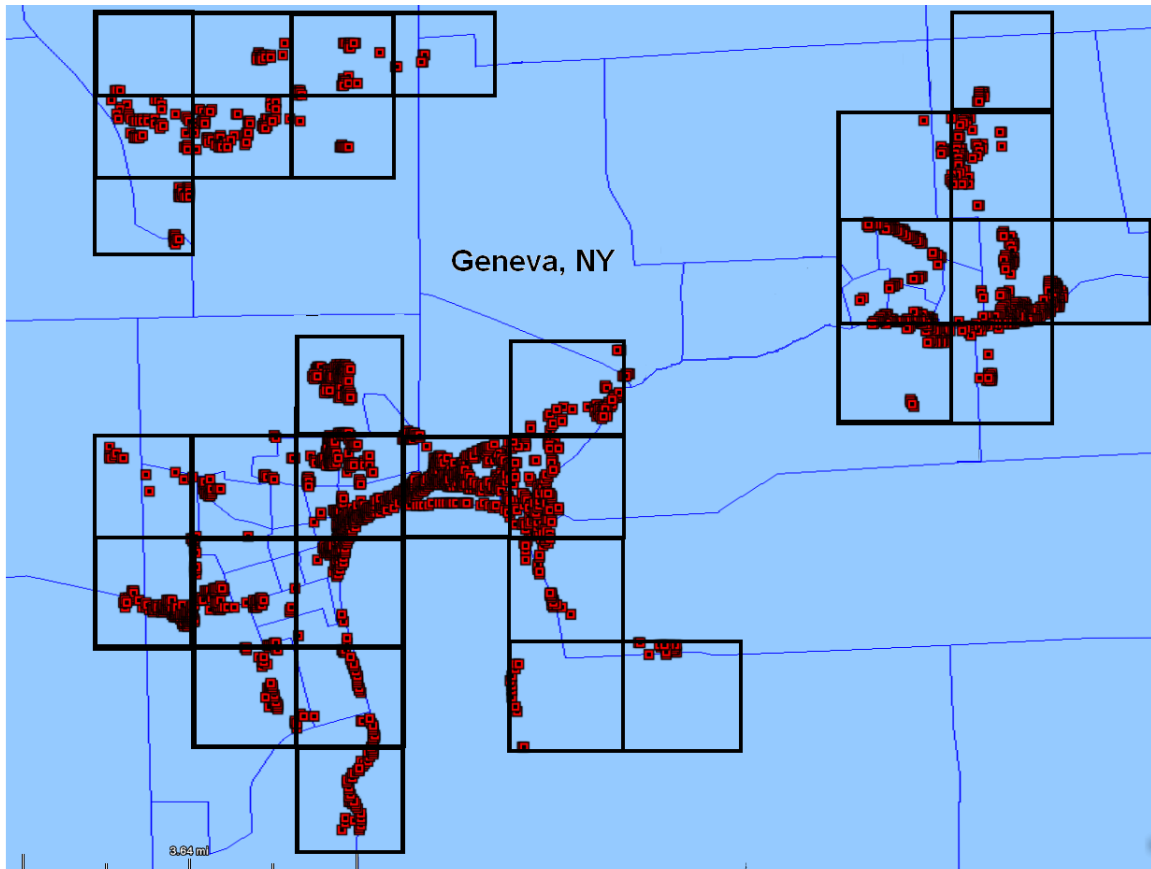


Figure 4g: Multiple Retail Zones within a market
(Geneva, NY; SM = Super Market Format; OT = Other Format)



Figure 4h: The pixel data covers retail businesses such as grocery stores, leaving out neighboring residential land (Showing NLCD pixel points over a 'Save-A-Lot' grocery store in Elmira, NY. Also, note that there is one pixel point for every 30mX30m land area.)

Figures 5: Co-location under different scenarios (x-axis: distance to nearest competitor in miles; y-axis: # stores; Percentage figures show the proportion of stores that are < 1mi. from a competitor)

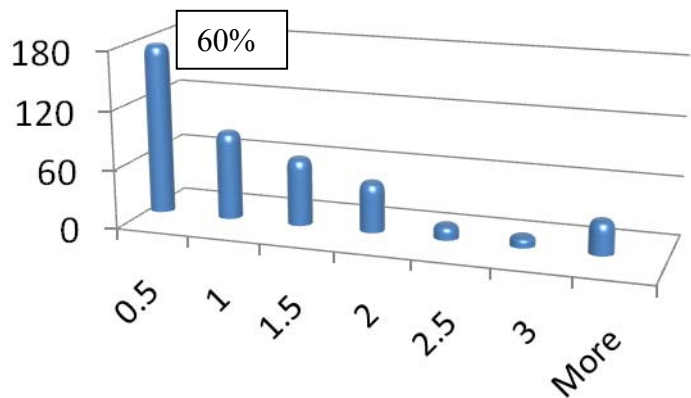


Figure 5a: Actual Locations (with Zoning + Agglomeration)

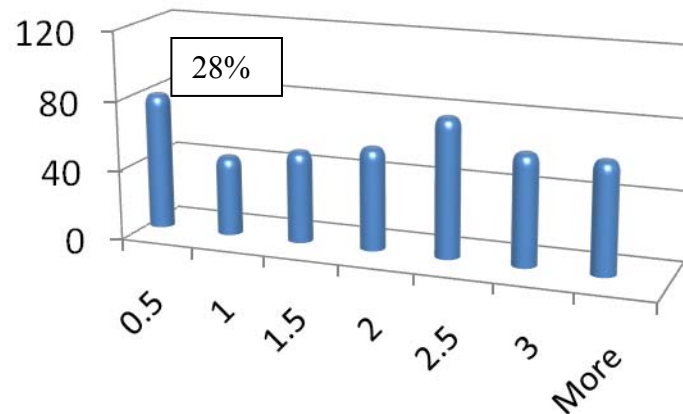


Figure 5b: Neither Zoning nor Agglomeration

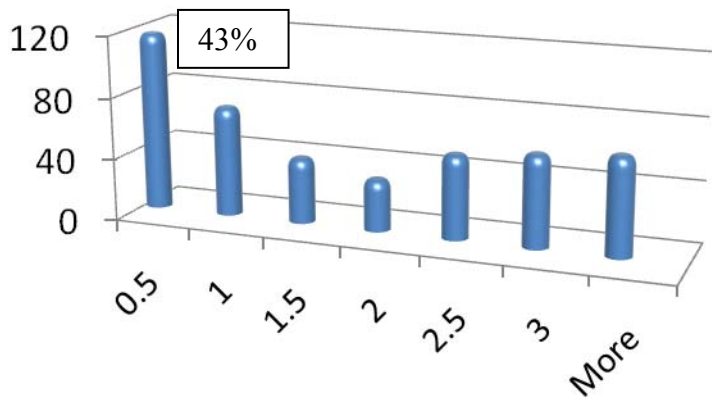


Figure 5c: Only Agglomeration

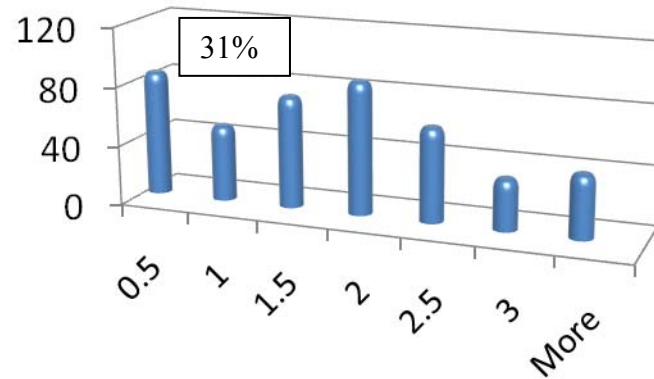
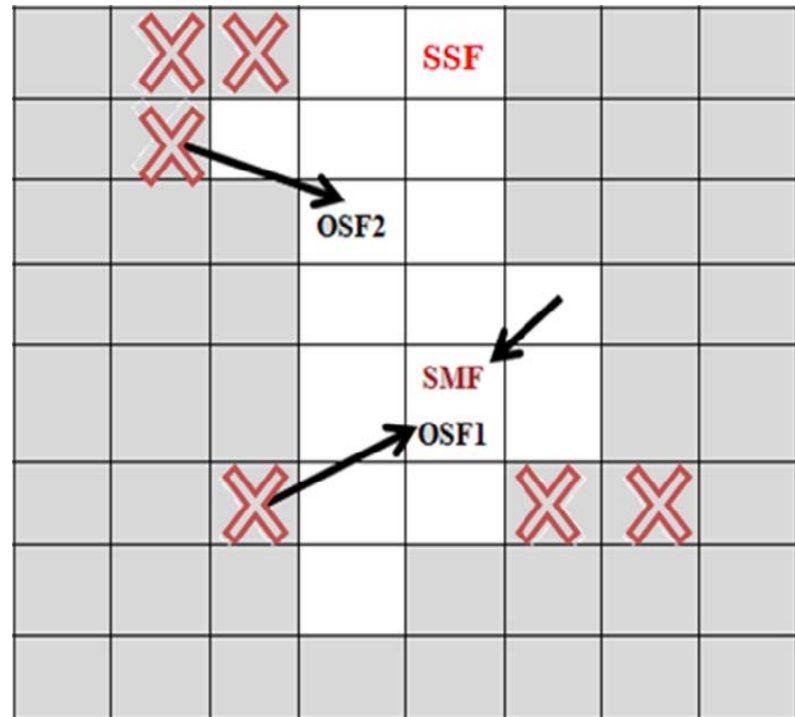
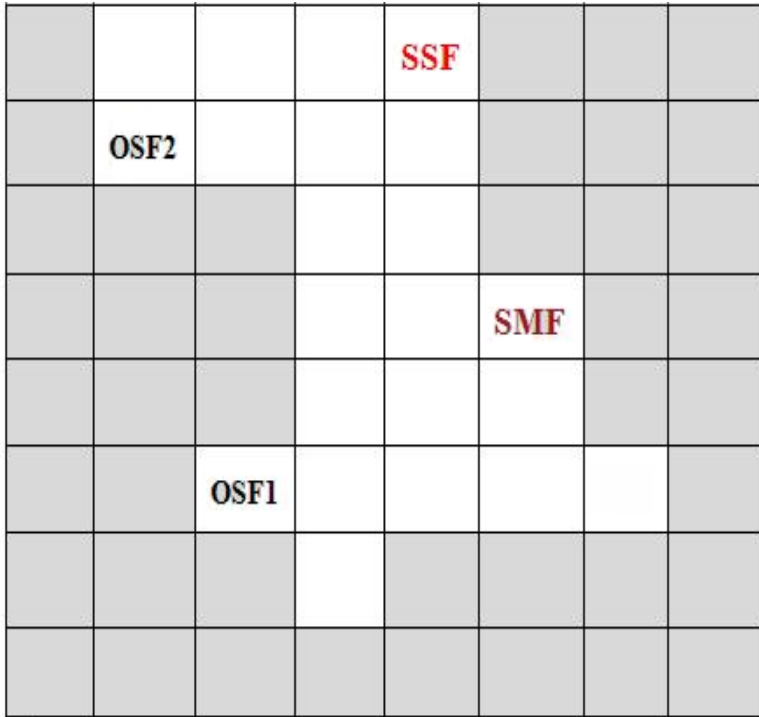


Figure 5d: Only Zoning



Notes: SMF – Supermarket format; SSF – Superstore format; OSF - Other store format.

Figure 6: Equilibrium Store Locations in a Simulated Market – Shrinking a Retail Zone (White region is a Retail zone)

Table 1: New Store Openings for Retailers

SPECIALITY	2001	2002	2003	2004	2005	2006	2007 (proj.)
Abercrombie	137	106	103	88	63	104	95
American Eagle	62	28	52	41	23	56	40
Ann Taylor	60	46	64	90	86	76	70
Bombay Company	11	3	49	31	-4	-46	-42
Buckle	21	9	12	11	11	16	16
Claire's Stores	-144	48	28	31	66	90	85
Coach	19	23	20	18	25	29	40
GAP	249	20	-95	-28	59	-3	0
Pacific Sunwear	129	73	86	113	115	100	95
Sharper Image	12	18	22	26	15	0	0
Talbot's	78	86	91	72	34	67	50
Total	634	460	432	493	493	489	449
ANCHOR							
Dillard's	1	-5	-5	1	1	5	5
Federated	18	-3	4	0	486	-80	5
JCPenney	-36	-14	17	1	-60	21	40
May's	18	4	1	57	-501	0	0
Nordstrom	15	11	13	2	7	2	5
Total	16	-7	30	61	-67	-52	55
COMMUNITY							
Cost Plus	48	0	29	33	33	17	23
Dollar General	540	573	587	620	609	525	-175
Dress Barn	31	34	18	4	-65	67	44
Family Dollar	452	475	411	439	432	275	335
Men's Wearhouse	29	9	41	4	18	25	5
Office Depot	-29	8	33	69	78	108	125
Pier 1	75	100	105	79	42	-25	-10
Ross Stores	43	55	61	81	71	75	90
TJX	172	178	219	162	161	101	104
Total	1,361	1,432	1,467	1,501	1,379	1,168	541
COMMUNITY/POWER ANCHORS							
Target	74	94	78	-245	89	98	115
Wal-Mart	315	364	372	349	352	349	260
Total	389	458	450	104	441	447	375
POWER ANCHORS							
BJ's	12	10	10	7	8	10	13
Bed Bath & Beyond	80	130	103	92	88	81	90
Circuit City	-5	2	-22	13	18	20	0
Costco	52	9	23	20	16	22	30
Home Depot	203	195	175	183	152	118	90
Lowe's	94	110	98	135	147	155	151
Total	436	456	387	450	429	406	374

Source: Company Reports, Merrill Lynch

Table 2: Average demand characteristics of locations with and without agglomeration

		Store Locations Where Nearest Rival Is More than 2 Miles Away	Store Locations Where Nearest Rival Is Within Half Mile
Population ($\times 10^3$)	0–1 mile	31.47	22.52
	1–3 miles	158.87	95.39
	> 3 miles	209.8	259.69
Per capita income ($\times 10^3$)	0–1 mile	19.4	20.8
	1–3 miles	18.9	20.0
	> 3 miles	20.5	19.8
Cumulative income ($\times 10^6$)	0–1 mile	57.39	45.00
	1–3 miles	280.12	185.73
	> 3 miles	430.80	513.83
# Retail businesses	0–1 mile	97.4	121.15
	1–3 miles	90.88	111.58
	> 3 miles	87.57	74.44

Table 3: Demand or Consumer side estimates

Variable	Parameter	Estimates
Travel Cost (T)		
Distance; d_{il}	α_1	0.1642 (0.1093)
Distance ² ; d_{il}^2	α_2	-1.2243 (0.2719)
Median HH Income* Distance; $med_hhl * d_{il}$	α_3	-0.0352 (0.0437)
Median Age*Distance; $med_age * d_{il}$	α_4	-0.0420 (0.0448)
Minimum Distance* Distance; $min_d * d_{il}$	α_5	0.1025 (0.0532)
Intrinsic Preference (λ)		
SSF_intercept	λ_0^{SSF}	-0.4517 (0.1324)
SSF_Median HH Income	λ_1^{SSF}	-0.1243 (0.0841)
SSF_Median Age	λ_2^{SSF}	-0.0133 (0.0316)
SSF_Minimum Distance	λ_3^{SSF}	0.1651 (0.0811)
OSF_intercept	λ_0^{OSF}	0.3145 (0.1072)
OSF_Median HH Income	λ_1^{OSF}	-0.2419 (0.1796)
OSF_Median Age	λ_2^{OSF}	0.0287 (0.0561)
OSF_Minimum Distance	λ_3^{OSF}	-0.1031 (0.0955)
Price	γ	-0.4683 (0.2513)
Economies of Scope		
Business Density; bus_l	α_6	1.9497 (0.3084)
Business density ² ; bus_l^2	α_7	-0.0326 (0.0713)
Store Agglomeration; N_l	α_8	0.7719 (0.2382)
Format Agglomeration; I_l^{MF}	α_9	1.2881 (0.3078)
Revenue Potential		
GSP_SMF; $GSP_{SMF,l}$	δ^{SMF}	0.4174 (0.1113)
GSP_SSF; $GSP_{SSF,l}$	δ^{SSF}	0.4423 (0.1928)
GSP_OSF; $GSP_{OSF,l}$	δ^{OSF}	0.4872 (0.1263)

Table 4: Supply or firm side estimates

Variable	Parameter	Estimates
SMF	τ_{SMF}	4.3814 (1.2418)
SSF	τ_{SSF}	4.1921 (0.9218)
OSF	τ_{OSF}	5.2186 (1.1534)
Intra-format competition		
SMF – SMF (0-2 mi.)	$\beta_{SMF-SMF,1}$	-1.2011 (0.1585)
SMF – SMF (2-4 mi.)	$\beta_{SMF-SMF,2}$	-0.5822 (0.1816)
SSF – SSF (0-2 mi.)	$\beta_{SSF-SSF,1}$	-1.6542 (0.1923)
SSF – SSF (2-4 mi.)	$\beta_{SSF-SSF,2}$	-0.8128 (0.2011)
OSF – OSF (0-2 mi.)	$\beta_{OSF-OSF,1}$	-0.9982 (0.1285)
OSF – OSF (2-4 mi.)	$\beta_{OSF-OSF,2}$	-0.3861 (0.2462)
Inter-format competition		
SMF – SSF (0-2 mi.)	$\beta_{SMF-SSF,1}$	-0.7326 (0.2445)
SMF – SSF (2-4 mi.)	$\beta_{SMF-SSF,2}$	-0.1623 (0.1098)
SMF – OSF (0-2 mi.)	$\beta_{SMF-OSF,1}$	-0.5611 (0.2165)
SMF – OSF (2-4 mi.)	$\beta_{SMF-OSF,2}$	-0.0264 (0.1328)
SSF – OSF (0-2 mi.)	$\beta_{SSF-OSF,1}$	-0.5271 (0.2422)
SSF – OSF (2-4 mi.)	$\beta_{SSF-OSF,2}$	-0.1008 (0.1028)
Unobserved shocks		
Price shock	ζ	1.8423 (0.7946)
Demand shock	ρ	1.1213 (0.2894)
Cost correlation	c	0.9546 (0.3018)
Retail zone scaling factor		
Mean	μ	-4.9217 (0.7149)
Std deviation	σ	1.1218 (0.2236)