An Empirical Model of Firm Entry with Endogenous Product-Type Choices

Katja Seim

Stanford Graduate School of Business, Stanford CA 94304[*]

September 17, 2004

Abstract

This paper presents a model of entry with endogenous product-type choices. These choices are formalized as the outcomes of a game of incomplete information in which rivals' differentiated products have non-uniform competitive effects on firms' profits. The model is estimated for location choices in the video retail industry using a nested fixed-point algorithm solution. The results imply significant payoffs to product differentiation. Simulations illustrate the tradeoff between demand and intensified competition and the extent to which markets with larger product spaces, and thus more scope for differentiation, support greater entry.

Keywords: Spatial differentiation, location choice, entry, retail markets

JEL Classification: L0, L1, L2, L8

[*]Email: kseim@stanford.edu. This paper is based upon various chapters of my Yale University Ph.D. thesis. I am grateful to Chris Timmins, Steve Berry, and Pat Bayer for their advice and help. Thanks go to Peter Davis, Ariel Pakes, John Parapatt, Ben Polak, Peter Reiss, John Rust and the participants at various seminars for their thoughtful comments.

1 Introduction

Firms frequently compete in the characteristics of products that they offer to consumers. Economic theory going back to Hotelling (1929) and Salop (1979) has analyzed such product differentiation as competition between products located at various positions in an abstract characteristics space, with consumers having idiosyncratic preferences over the different positions. A firm's choice of location in the characteristics space becomes a strategic variable, depending endogenously on the positioning choices of its competitors.

While previous studies of entry have modeled the trade-off between the available demand and the intensity of competition faced by a new entrant, the strategic importance of product positioning within a market has received less attention. Product differentiation allows a firm to both better serve consumers' differing preferences and potentially acquire localized market power. One possible determinant of the extent of such localized market power is the scope for differentiation afforded by the product. For example, is localized market power and thus the incentive for entry greater for complex products like automobiles with vast differentiation opportunities than for products such as yogurt that are characterized by fewer attributes? Or does the larger diversity of consumer preferences associated with more complex products erode such profit opportunities due to a lack of sufficient demand for highly differentiated varieties?

This paper presents an empirically tractable equilibrium model to analyze such determinants of firms' product positioning decisions. It can also be extended to a number of economic contexts. Work by Augereau, Greenstein, and Rysman (2004) analyzes the importance of competition in a technology adoption game between Internet Service Providers and suggests differentiation between providers as a possible explanation for the lack of a single technological standard being chosen. Watson (2004) uses a similar model to study how firms' product variety choices depend on separation from rivals in geographically-differentiated markets and finds that higher localized competition promotes increased variety in the context of eye-glass retailing. In general, the framework can easily accommodate a multitude of horizontal or vertical product differentiation choices, including discrete product attributes or combinations of attributes or continuous attributes that can be discretized into distinct categories.[1]

[1] For example, Mazzeo (2002) looks at quality levels in the motel industry by discretizing a continuous range into three quality tiers.

The empirical results for the sample of video retailers bear out the intuition that firms use spatial differentiation to shield themselves from a large fraction of competitors in the market. Distant competitors drive down payoffs significantly less than competitors close-by and such competitive interaction helps to explain the patterns of location choices found in the data. The significant payoffs to product differentiation motivate an analysis of the role of the size of the characteristics space, measured simply by the market's area in this retail application, in affecting market structure. I illustrate the effect on the entry process of the geographic dispersion of demand as a market's area and population grow. As the market area grows, the ability of firms to achieve greater localized market power through the increased scope for spatial differentiation is found to have a significant impact on overall entry. At the same time, however, payoffs from such differentiation are offset by the negative effect of reduced demand due to the dispersion of population in the increased market area. I find that for the specific example of the video retail industry with highly localized demand, the net effect of these two forces is that as a market's population and area grow, the number of firms supported by the market increases only slightly. The example illustrates that the extent of localized market power generated by product differentiation is the outcome of the interplay between differentiation possibilities and the strength of consumer preferences for the particular product attribute.

The model presented addresses several difficulties in the empirical implementation of models of product differentiation and market structure. First, previous work on extending empirical models of market structure to product space positioning yields equilibrium location strategies that are tractable only for a severely limited number of product positions that is unrealistically small in many applications. Second, the estimation of demand at various product space positions requires the use of detailed data on prices and quantities sold across all competitors in the market. Price and, in particular, quantity data is frequently not available. Last, a general difficulty in estimating discrete strategic games is that the interdependence of players' choices oftentimes entails the existence of multiple equilibria.

The first difficulty in using discrete games to model product positioning decisions is that it is computationally difficult to verify a Nash equilibrium. The market structure literature (e.g. Bresnahan and Reiss 1988, Bresnahan and Reiss 1991, Berry 1992) sets up firms' entry decisions as the equilibrium outcome of a discrete game played between potential entrants. Mazzeo (2002) extends this framework to a model of discrete product space positioning that predicts firms' joint entry and quality-level choice.[2] An analytically appealing feature of his model is that firms possess complete information about competitors. A given firm distribution across locations constitutes an equilibrium if each firm maximizes profits and none of the competitors has an incentive to deviate after having made its location choice. To confirm that a given firm configuration is an equilibrium entails verifying that the specific configuration is more profitable for each and every firm than any other possible configuration. A drawback to the model is that computing an equilibrium configuration is difficult for markets with large numbers of locations and firms. The application of the model to the motel industry shows that even with three quality levels, estimation becomes burdensome due to the large number of profit constraints that must hold.

In contrast to the complete information framework, the model presented here recognizes that idiosyncratic sources of profitability, which are difficult to observe by competitors, may be of importance in firm behavior. Examples include differences in firms' cost structures or their intangible assets, such as staff managerial talent, customer service, and inventory maintenance. While such factors affect firm profitability, they cannot be easily observed by competitors who must separate the role of such intangible assets in a firm's success from other related profit shifters. I allow firms to possess private information about their own profitability, which is observed by competitors only in distribution. Payoffs are thus a function of the firm's expectation of its competitors' optimal location choices, as well as its own idiosyncratic profitability. As Rust (1994) recognizes, the resulting Bayesian Nash equilibrium conjectures, which represent the probability of adopting a particular strategy, can be derived more easily than in the complete information framework, which solves for an exact equilibrium strategy. This set-up facilitates the application of the model to a realistic, large-dimensional set of product types. This is only one avenue where the use of asymmetric information between players simplifies the computation of equilibrium strategies. Recent work on the estimation of discrete dynamic games (e.g. Aguirregabiria and Mira 2004, Pakes, Ostrowsky, and Berry 2004, Pesendorfer and Schmidt-Dengler 2003) similarly draws upon information asymmetries between players to reduce the burden of computing dynamic entry and exit strategies.

To overcome the second, data-driven, hurdle, previous models (e.g. Sutton 1991, Sutton 1998) assume that market structure reveals information about firms' underlying economic profits, which in turn provides information about the intensity of competition between firms. In the absence

[2] Related work in this area includes Ellickson (2004), Stavins (1995), and Toivanen and Waterson (2001b). For reviews of empirical research on discrete choice game-theoretic models, see Toivanen and Waterson (2001a) and Reiss (1996).

of price and quantity data, more easily observable entry decisions can then be used to infer firm profitability. Building upon this concept, Bresnahan and Reiss (1988, 1991) and Berry (1992) model entry as a function of demand proxies and the competitive intensity due to a market's attractiveness to rivals. Since the authors consider well-defined, homogeneous products and markets, the firms in their studies have no scope to offer customers differentiated products. Due to a lack of detailed outlet pricing and rental data, I similarly exploit information revealed by firms' location choices as an indicator of profitability.

The last difficulty in estimating discrete strategic games relates to the multiplicity of equilibria generated by earlier models. The majority of the existing literature (e.g. Berry 1992, Bresnahan and Reiss 1991, Berry and Waldfogel 1999) specifies models that combine multiple equilibria for individual players' actions into one joint equilibrium that can be uniquely predicted.[3] In the entry context, for example, such models uniquely predict the equilibrium number of entrants into a market, but not the identity of such entrants. The assumption of an asymmetric information structure embedded into a discrete game does not solve the problem of multiplicity of equilibria in general. I show, however, that for a simplified version of the model studied here, a unique equilibrium exists under reasonable restrictions on the model parameters. I present simulation evidence for the more general version of the model.

The remainder of the paper proceeds as follows. Section 2 presents the model of entry and location choice and discusses the econometric estimation of the model. Section 3 then describes the data set used for estimation, while section 4 presents the results of the estimation and counterfactual analyses of the contribution of spatial differentiation to the entry process. Section 5 concludes with a discussion of the main features of the model and their economic implications.

[3] Recent work by Tamer (2003), Andrews, Berry, and Jia (2004), and Ciliberto and Tamer (2004) proposes instead to derive an upper and a lower bound for the probability of observing each non-unique outcome of the original game. The estimation similarly allows for the identification of bounds on the parameters of interest only rather than point estimates. Ciliberto and Tamer (2004) also present an empirical application of these procedures to airline route selection.

2 Model

2.1 Set-up and Payoffs

The model treats a firm's strategy as a choice problem of selecting the optimal product position among a set of possible discrete locations in characteristics space. To ensure that equilibrium strategies can be easily computed for many locations, the model employs a static set-up. Accordingly, each firm makes its joint entry and location choice based on a comparison of expected post-entry, single-period profits across locations.

Two opposing forces drive profits. On the one hand, favorable demand characteristics attract firms to a location. At the same time, the increased concentration of firms at that location adversely impacts firm profits due to greater competition. Following Manski (1993), firm behavior is endogenously determined by the choices of other firms. Profits also vary between firms in the same location due to differences in cost structures and other firm-specific sources of profitability. By assumption, such differences are only observed by the firm itself. Entry and location choices are thus determined by the demand characteristics in each market location, firms' expectations of their competitors' optimal location choices, and each firm's idiosyncratic profitability.

Formally, a set of $\mathcal{F}$ potential entrants simultaneously chooses whether or not to enter a market m and where in m to locate. The number of potential entrants is known to all firms and exceeds one. For simplicity, firms are assumed to make independent entry and location choices.[4] The number of actual entrants into the market is denoted by $\mathcal{E}$. The set of possible locations in the market is indexed by $l = 0, 1, ..., \mathcal{L}^m$, which includes the decision not to enter given by $l = 0$. Firm f's location decision, with $f = 1, ..., \mathcal{F}$, is denoted by $\boldsymbol{d}_f$, where $d_{fl} = 1$ if location l is chosen, 0 otherwise.

[4] The model thus ignores one dimension of the decision-making process for multi-product firms and chain stores, namely the potential for cannibalization of revenues of existing products or established stores when introducing a new product or opening a new outlet. The significance of this assumption will depend in each case upon the particular empirical setting. I discuss the importance of chain stores in the sample markets further in section 3.

Upon entry, firm f's payoff in location l in market m is given by the following reduced form:[5]

$$\begin{aligned}\Pi_{fl}^{m} &= \boldsymbol{X}_l^m \boldsymbol{\beta} + \xi^m + h(\Gamma_{.l}^m, \boldsymbol{n}^m) + \varepsilon_{fl}^m \\ &= \overline{\Pi}_l^m + \varepsilon_{fl}^m.\end{aligned} \tag{2.1}$$

The first two terms represent demand characteristics that affect payoffs in location l in market m. $\boldsymbol{X}_l^m$ is a vector of demand and cost characteristics specific to location l, such as population or income. Since the demographic characteristics that can be observed by the econometrician may not reflect all cost and demand factors driving firm profitability, unobservable exogenous differences across markets are captured by a market-level characteristic, ξ^m. All cost and demand shifters are known to the firm and its competitors at the time of decision-making. The next term, $h(\Gamma_{.l}^m, \boldsymbol{n}^m)$, captures the effect on profits due to competition from all rivals in the market. Non-uniform competitive interaction between different product types requires, in the case of a two-dimensional characteristics space, a matrix of competitive effects by location pairs. Therefore, Γ is an $\mathcal{L}^m \times \mathcal{L}^m$ matrix of competitive effects; for example, the lth column of Γ, $\Gamma_{.l}^m$, represents the competitive intensity between competitors in locations 1 through $\mathcal{L}^m$ and a firm in location l. The elements of Γ vary by the similarity between products, that is their distance in characteristics space. For the case of spatial differentiation, the appropriate distance metric that measures the similarity between firms' products is simply physical distance. The impact on payoffs due to competition from other firms is thus a function of $\Gamma_{.l}^m$ and $\boldsymbol{n}^m$, where $\boldsymbol{n}^m$ is a vector containing the number of firms in each of the $\mathcal{L}^m$ locations in the market. Mean profits from not entering, $\overline{\Pi}_0$, are normalized to zero across firms and markets.

ε_{fl}^m represents the idiosyncratic component of firm f's profits from operating in location l. As in Rust (1994), the asymmetry of information between firms arises from this idiosyncratic profitability (their "type"), which is treated as a realization of a random variable whose distribution is common knowledge among all competitors, but whose value is private information to the firm. Players' information sets and types are defined by the following assumption:

[5] This specification of the profit function is chosen primarily to overcome the unavailability of firm-level market shares. This limitation on available data is shared by other work in the field, such as Berry (1992) and Bresnahan and Reiss (1991). To model homogeneous product markets, they use a similar payoff function interpreted as the log of a market size/demand term multiplied by a variable profit term that depends on the number of market competitors. Correspondingly, the video rental market setting in this paper could be viewed as offering a largely homogeneous product (the video itself), while store characteristics, and in particular stores locations, are the main factors of differentiation between firms. For an example of a study that uses firm-level market share information, see Berry and Waldfogel (1999).

(A1) *Independent symmetric private values:* Players' profitability types ε_1^m, ...,$\varepsilon_{\mathcal{F}}^m$ are private information to the players and are independently distributed draws from the same prior distribution $G(\cdot)$.

In this specification, ε, a firm's type, captures *all* differences between it and other potential entrants. The payoff function thus retains some of the symmetry underlying the payoff functions common to the previous literature. Profits depend only on the number of entrants at every location, and not on the entrants' identities.[6] The symmetry assumption implies that each pair of firms will have the same conjecture about the profitability of a third firm and the profitability of any pair of firms is identically distributed.

For the purposes of estimation, I make the following further assumptions, which allow an identical profit function to be applied to every location in the market and across markets with varying numbers of locations:

(A2) $h(\Gamma_{.l}^m; \boldsymbol{n}^m) = \sum_{k=1}^{\mathcal{L}^m} \gamma_{kl} n_k^m$.

(A3) $\gamma_{kl} = \gamma_{k'l} = \gamma_b$ if $d_b \leq d_{kl}^m, d_{k'l}^m < d_{b+1}$, where d_b and d_{b+1} denote cutoffs that define a distance band.

Assumption (A2) implies that competitors' effects are additively separable across locations and that the incremental impact on payoffs of an additional firm in a given location is constant. Assumption (A3) accommodates irregularities in the data that affect estimation of the model. The sample market locations, as further described in Section 3, consist of population-weighted centroids of Census tracts. Census tracts vary in area and shape due to differences in population between tracts as well as regional differences in city planning. The irregularity of Census tracts implies that no two-tract centroids are at the exact same distance from each other as others in the market. Assumption (A3) implies that firms located in different cells, k and k', but within a given distance range from location l have the same impact on the profitability of firms in location l, across locations in a market and across markets.[7] As a result, rivals located in tracts within a certain distance band around location l exert the same competitive pressure on a firm in location l.

[6]Symmetry is maintained in this paper solely for computational reasons. It can be relaxed to allow for more heterogeneous payoff functions. Einav (2003), for example, estimates a sequential, Bayesian timing game applied to heterogeneous movie producers' choices of movie release dates.

[7]Note some of the potential problems with this approach: the competitive pressure exerted by two firms located in a single location to the north of l will be the same as the competitive pressure of two firms, of which one operates in the cell directly north of l and one in the cell directly south of l. Ideally, the first scenario would be more attractive to firm f than the second; however, in the current treatment of competitive impacts, there will be no difference.

Allowing for a maximum of B distance bands, indexed by $b = 0, 1, ..., B$, the resulting payoff function is given by

$$\Pi_{fl} \quad = \quad \xi + \boldsymbol{X}_l \boldsymbol{\beta} + \sum_b \gamma_b N_{bl} + \varepsilon_{fl}, \tag{2.2}$$

where market superscripts m have been omitted to simplify the exposition. In equation 2.2, γ_b represents the impact of competitors in distance band b. γ_0 measures the competitive effect of firms at a distance between zero and d_1, γ_1 the competitive effect of firms at a distance between d_1 and d_2, and so forth. The aggregate number of firms in locations within each of these distance bands is given by N_b. Specifically, $N_{bl} = \sum_k \mathbb{I}_{kl}^b n_k$, where $\mathbb{I}_{kl}^b = 1$ if $d_b \leq d_{kl} < d_{b+1}$ and 0 otherwise. Summing N_b across distance bands yields the total number of competitors in the market, $\mathcal{E}$.

2.2 Conjectures and Equilibrium

The location decision has to account for the fact that due to imperfect information about its rivals' profitability, a firm can only form an expectation of their optimal location choices. Based on this expected competitor distribution across market locations, each firm will choose the location that maximizes its payoffs given its realization of its own profitability type. The expected profit in location l is

$$\mathbb{E}[\Pi_{fl}] = \xi + \boldsymbol{X}_l \boldsymbol{\beta} + \sum_b \gamma_b \mathbb{E}[N_{bl}] + \varepsilon_{fl}, \tag{2.3}$$

where the expected number of firms per distance band, $\mathbb{E}[N_{bl}]$, now equals $\sum_k \mathbb{I}_{kl}^b \mathbb{E}[n_k]$.

Assumption (A2) that competitors impact firm profits linearly greatly simplifies the computation of the expected competition from all firms. The linear specification of competitive interaction is in contrast to previous work by Berry (1992) and Bresnahan and Reiss (1991) who employ more flexible functional forms for $h(.)$ that decrease in $\boldsymbol{n}$ at a declining rather than constant rate. Relaxing assumption (A2) to incorporate a more desirable functional form for $h(\gamma, n)$ is of computational rather than conceptual difficulty as it would involve the use of more complicated numerical integration techniques.

Due to the symmetry of rivals' types, firm f's probability assessment of firm g's optimal location strategy is the same across all competitors. The probability that competitor g chooses location l, p_{gl}, is given by

$$p_{gl}(d_{gl} = 1 | \xi, \boldsymbol{X}, \mathcal{E}, \theta_1) = \Pr(\mathbb{E}[\overline{\Pi}_{gl}(\cdot)] + \varepsilon_{gl} \geq \mathbb{E}[\overline{\Pi}_{gk}(\cdot)] + \varepsilon_{gk}, \forall\ k \neq l,\ \forall\ g \neq f), \tag{2.4}$$

combining the payoff function parameters $(\boldsymbol{\beta}, \gamma)$ into θ_1. The number of competitors that firm f expects to face in location l then equals $(\mathcal{E}-1)p_{gl}$ and the expected number of firms entering each of the distance bands b collapses to

$$\mathbb{E}[N_{bl}] \;=\; \sum_k \mathbb{I}^b_{kl}\mathbb{E}[n_k] = \sum_k \mathbb{I}^b_{kl}(\mathcal{E}-1)p_{gk} + \mathbb{I}_{b=0}. \tag{2.5}$$

The indicator variable $\mathbb{I}_{b=0}$, set equal to one for distance band $b = 0$ and zero for all remaining distance bands, reflects that the number of firms in distance band $b = 0$ includes firm f itself were it to choose location l.

For the sake of computational tractability, I assume that players' types, ε, are i.i.d. draws from a type 1 extreme value distribution.[8] The scale parameter of the extreme value distribution captures the degree of uncertainty that a firm has over its rivals' profitability draws. In the limit, as it tends to zero, profitability draws across locations are perfectly correlated and concentrated at the distribution mean. In this case, there is no uncertainty about rivals' profitability and McKelvey and Palfrey (1995) show that the outcome of the game approaches that of the corresponding perfect information model. In the empirical estimation, the scale parameter is not separately identified from the remaining parameters of the payoff function and is therefore normalized to one. This results in traditional multinomial Logit probabilities for firms' beliefs, conditional on the entry of $\mathcal{E}$ firms, given by

$$p_{gl} = \frac{\exp(\mathbb{E}[\overline{\Pi}_{gl}])}{\sum_{k=1}^{\mathcal{L}} \exp(\mathbb{E}[\overline{\Pi}_{gk}])}. \tag{2.6}$$

As an illustration, figure 1 depicts a square-shaped market made up of nine locations, which are grouped into three distance bands. Consider the payoffs to locating in cell seven. Immediate competitors are specified to be those in the firm's own location. The neighboring competitors, namely those in band 1, of a firm that chooses this location are the firms located in cells four, five,

[8] As suggested by Rust (1994), the extreme-value specification is attractive in this context because it entails closed-form expressions for players' choice probabilities. The computational tractability of i.i.d. Logit draws comes, however, at a cost. It implies that firm profitability is uncorrelated across *firms* within a given location, as well as across *locations* for a given firm. Thus the specification does not consider that profitability is likely to exhibit spatial correlation since demand characteristics are spatially correlated. If such patterns were of importance, the estimated competitive effects would be biased downwards. To allow for more realistic substitution patterns across locations, random parameter coefficients or a more flexible error distribution could be used. Most appropriate distributions, such as the multinomial normal distribution, do not yield closed-form solutions for the vector of location probabilities, and in either case, the computational complexity of finding equilibrium probabilities would increase significantly.

and eight, while the most distant competitors are located in cells one, two, three, six, and nine. Based on equation 2.3, $\mathbb{E}[\overline{\Pi}_7]$ equals

$$\mathbb{E}[\overline{\Pi}_7] = \xi + \boldsymbol{X}_7\boldsymbol{\beta} + \gamma_0 + (\mathcal{E}-1)(\gamma_0 p_7 + \gamma_1(p_4 + p_5 + p_8) + \gamma_2(p_1 + p_2 + p_3 + p_6 + p_9)). \tag{2.7}$$

Assuming that the competitive impact of neighboring firms, γ_1, exceeds the impact of more distant firms, γ_2, the appeal of cell seven lies primarily in its placement at the edge of the city with a small set of immediately adjacent locations and competitors. In contrast, a firm located in cell five will have many close-by competitors, exposing it to stronger competition than a firm located on the city's fringe. At the same time, from a demand perspective, cell five is more attractive than cell seven because it grants easy access to most of the consumers in the market living in its own and neighboring locations. The equilibrium firm location pattern is then determined by this trade-off between demand and competitive pressures.

The equilibrium is a symmetric Bayesian Nash equilibrium describing the optimal response that maximizes the firm's expected payoff conditional on entry, given its conjecture about other competitors' strategies. The assumption of independent symmetric types implies that every firm has the same equilibrium conjecture of its competitors' location choices, namely $\boldsymbol{p}_g$=$\boldsymbol{p}_f$=$\boldsymbol{p}^*$. A firm's vector of equilibrium conjectures over all locations l is then defined recursively by the following set of $\mathcal{L}$ equations

$$p_l^* = \frac{\exp(\overline{\Pi}_l(\boldsymbol{X}, \boldsymbol{p}^*, \mathcal{E}, \theta_1))}{\sum_{k=1}^{\mathcal{L}} \exp(\overline{\Pi}_k(\boldsymbol{X}, \boldsymbol{p}^*, \mathcal{E}, \theta_1))} = \frac{\exp(\boldsymbol{X}_l\boldsymbol{\beta} + \gamma_0 + (\mathcal{E}-1)\sum_b \gamma_b \sum_j \mathbb{I}_{jl}^b p_j^*)}{\sum_{k=1}^{\mathcal{L}} \exp(\boldsymbol{X}_k\boldsymbol{\beta} + \gamma_0 + (\mathcal{E}-1)\sum_b \gamma_b \sum_j \mathbb{I}_{jk}^b p_j^*)} \tag{2.8}$$

$$\forall\ l = 1, ..., \mathcal{L}$$

where the expressions for the expected number of firms per distance band have been substituted into the payoff function. This system of $\mathcal{L}$ equations defines the equilibrium location conjectures as a fixed point of the mapping from the firm's conjecture of its competitors' strategies into its competitors' conjectures of the firm's own strategy. The Appendix describes the existence and uniqueness properties of the location equilibrium for the specific payoff function used in estimation.

In equation 2.5, a firm's expected number of competitors in a particular distance band is a function of the number of entrants into a market. In equilibrium, each entrant earns non-negative profits in expectation, while any additional entrant would suffer losses. The probability of entry by a firm into a market involves a comparison of a weighted average of payoffs across locations to the normalized payoff of not entering. Given the assumption of i.i.d. extreme value profitability types, the probability of entry is given by:

$$\Pr(\text{entry}) = \frac{\exp(\xi)\left[\sum_{l=1}^{\mathcal{L}} \exp(\overline{\Pi}_l(\boldsymbol{X}, \boldsymbol{p}^*, \mathcal{E}, \theta_1))\right]}{1+\exp(\xi)\left[\sum_{l=1}^{\mathcal{L}} \exp(\overline{\Pi}_l(\boldsymbol{X}, \boldsymbol{p}^*, \mathcal{E}, \theta_1))\right]}. \tag{2.9}$$

Note that while market-level factors do not affect the attractiveness of any one location differentially, they influence the firm's overall entry decision. Such market-level factors are captured in the payoff function in 2.3 by ξ. The probability of entry is identical across competitors and as a result, the expected number of entrants is simply

$$\mathcal{E} = \mathcal{F} \cdot \Pr(\text{entry}). \tag{2.10}$$

Through equation 2.9, the expected number of entrants depends on the equilibrium location conjectures, which in turn depend on the expected number of entrants. Solving for equilibrium location and entry probabilities thus requires knowledge of $\mathcal{F}$. Most available data sets only include information on actual entrants but not those that merely consider entering, but choose not to. An earlier study of entry into grocery retailing (Cotterill and Haller 1992) has dealt with this problem by setting the potential entrant pool equal to the number of major chains in the industry. Due to the significant presence of non-chain affiliated retailers in the current application, I instead estimate the model by fixing the potential entrant pool exogenously at different values. One alternative I consider is setting the number of potential entrants such that 50% eventually enter into the market. As a second alternative, I assume that a fixed number of firms consider entry into each market, equivalent to the approach used by Cotterill and Haller (1992). I use a potential entrant pool of 50 firms, which is for most markets significantly larger than twice the number of actual entrants, allowing for a comparison of the estimated parameter values under two very different assumptions on potential entrants.

To complete the game-theoretic model, the system of equations in 2.8 can be augmented by 2.10 to yield a joint equilibrium prediction for the location probabilities and the number of entrants. Since a closed-form solution for the equilibrium does not exist, it is found using the method of successive approximations from an initial guess of the equilibrium.

2.3 Estimation

The augmented system of equations 2.8 and 2.10 is highly non-linear, which in practice proves to be numerically difficult to solve repeatedly, in particular if the initial guess of the number of entrants into a market is far from the ultimate solution to equation 2.10. Instead, in the empirical implementation of the model, I simplify the treatment of the entry decision by assuming that the expected number of entrants predicted by the model in 2.10 perfectly equals the number of entrants observed in the data. To do so, I adjust the market-level effect ξ until the expected number of entrants defined in equation 2.10 equals the observed number of entrants. Equations 2.9 and 2.10 can be solved for ξ as a function of location characteristics, equilibrium location conjectures, and the number of potential and actual entrants:

$$\xi \quad = \quad \ln(\mathcal{E}) - \ln(\mathcal{F} - \mathcal{E}) - \ln\left(\sum_{l=1}^{\mathcal{L}} \exp(\overline{\Pi}_l(\boldsymbol{X}, \boldsymbol{p}^*, \mathcal{E}, \theta_1))\right). \tag{2.11}$$

Upon substitution of the observed number of entrants in a market for $\mathcal{E}$, equation 2.11 yields a market-specific realization for ξ. For this particular realization of ξ, the number of entrants predicted by the model coincides with the observed number of entrants in each market. This approach of using an unobservable effect to induce an equivalence between actual and predicted numbers of entrants follows the approach used by Berry (1994) and Berry, Levinsohn, and Pakes (1995) in their estimation of competition in differentiated product markets. To close the econometric model, I assume that ξ is a random effect that is distributed normally with mean μ and standard deviation σ, independently from the location-specific unobservable ε. Thus, the realization of ξ that results from equation 2.11 is treated as a draw from a Normal distribution whose mean and standard deviation are parameters to be estimated based on the vector of ξ across the set of M markets.

Estimation proceeds via maximum likelihood. Each market is treated as an independent $\mathcal{F}^m$-player location game. Based on the payoff function, I predict the discrete location choices of each of the $\mathcal{F}^m$ potential entrants in a market. The dependent variable consists of a vector of each firm f's observed location choice, stacked across firms and markets. The likelihood function is given by

$$L(\theta_1, \theta_2) = \prod_{m=1}^{M} \boldsymbol{p}_{\theta_1}(\boldsymbol{d}^m | \xi^m, \boldsymbol{X}^m, \mathcal{E}^m) g_{\theta_2}(\xi^m | \boldsymbol{X}^m, \mathcal{E}^m, \mathcal{F}^m), \tag{2.12}$$

where $\boldsymbol{d}^m = (d_1^m, d_2^m, ..., d_{\mathcal{F}}^m)$ denotes the vector of actions taken by the $\mathcal{F}$ players in market m. The likelihood function consists of two parts. The first part computes the likelihood of observing entrants' location choices conditional on the market-level effect ξ, that is the Logit location choice probabilities. To derive the unconditional likelihood, the location-choice probability is multiplied by the probability of observing the particular ξ realization that equates predicted and actual entrants. The Normal density of each observation ξ^m is denoted by g_{θ_2}, with $\theta_2 = (\mu, \sigma)$.

For a given set of values for the parameter vector (θ_1, θ_2) and data on location characteristics and actual and potential entry in each market, the system of equations in 2.8 is solved numerically for its fixed point on a market-by-market basis. Successive approximations to the fixed point result in a vector of equilibrium location choice probabilities, $\boldsymbol{p}^m$.[9] The equilibrium location choice probabilities, together with $\mathcal{E}$ and $\mathcal{F}$, feed into equation 2.11 to yield an equilibrium realization of the market-level unobservable ξ for each market m.

The fixed-point algorithm is then nested into a maximum likelihood routine to find the optimal parameters to explain the observed firm behavior. The parameters to be estimated include both $\boldsymbol{\beta}$ and γ, which characterize the payoff function, and the parameters describing the distribution of the market-level random effect, μ and σ. Parameter estimates are obtained by maximizing equation 2.12 using a Nelder-Meade optimization algorithm. Starting values for the optimization routine are found by performing a grid search over the parameter space.

The estimation procedure relies on matching a set of location probabilities to the actually observed location choices by the video retailers in the sample. This approach would be problematic if the location-choice game had multiple equilibria. The Appendix investigates the uniqueness properties of the model for payoff functions with two and three distance bands. I show that for the payoff function with two distance bands applied to a market with four locations only, the model has a unique equilibrium provided competitive intensity between firms decreases with distance. For the more general payoff function with three distance bands covering an arbitrary number of locations, a similar constraint on the parameter values cannot be established analytically since the complexity of the expressions for the probabilities increases in the number of locations and firm types. Simulation evidence suggests, however, that the result from the simpler model carries over. The numerical fixed point algorithm converges to a single solution either when the intensity of competition decreases

[9] Explicitly solving for the equilibrium may be computationally burdensome for more complex models than the one considered here. Alternatives to using a nested fixed-point algorithm have been suggested by Ahn and Manski (1993) for a binary choice model under uncertainty and by Aguirregabiria and Mira (2004) for a dynamic discrete game of imperfect information. These approaches rely on initial non-parametric estimates of the respective expectations and equilibrium choice probabilities to then derive players' optimal decisions and draw an inference of the underlying preferences.

with distance or when the market locations differ from each other in their characteristics, generating exogenous variation for competitors' expected location choices. Such variation is present in the sample locations' exogenous characteristics, as discussed in section 3.2. Consequently, I do not impose any restrictions on γ during the estimation, and I verify ex post that the equilibrium associated with the optimal parameter values is unique.

3 Data

The model is applied to entry and product-type choices in the video retail industry using discrete location choices as product space positions. The video retail industry is well suited for an analysis of location choice as an instrument of product differentiation. The transaction under consideration consists of the rental of a videotape, a homogeneous and relatively inexpensive good, with prices of the rental transaction ranging between \$2 to \$4 per tape. Since videotapes are standardized, stores differentiate themselves in other ways, including the variety and depth of inventory carried, the terms of the rental contract concerning the rental period, and drop-off convenience. The main avenue of differentiation arises, however, from spatial location since the small absolute differences in prices across stores make customers unwilling to travel a long distance to carry out the rental at a lower price. This spatial dimension of product differentiation is the main focus of the paper.[10]

3.1 Sample Markets

Spatial differentiation will only play a significant role in market structure if the population, or the available demand, is sufficiently large and geographically spread out for firms to strategically exploit location. According to research commissioned by the Video Software Dealers' Association (1998), the average customer travels only 3.2 miles for a round trip to a video store. The markets used in this study are selected, therefore, to provide adequate scope for spatial differentiation by firms, while not being so large that distant competitors would rarely, if ever, compete with each other for customers. To facilitate identification of competitors operating within each market as

[10] Seim (2001) contains an extension of the model that allows firms to differ along other observable dimensions by incorporating two types of firms, larger stores with a deeper inventory (proxied by whether stores are affiliated with a chain) and smaller stores that represent a different mix of product characteristics that has also found success in the marketplace. While I find that competition differs between stores of different storefront types, the analysis is significantly more complex and is therefore omitted from this paper.

well as potential customers in the market, I focus on well-delimited cities or groups of cities with shared boundaries. Starting from Census data on medium-sized cities or incorporated places with a population between 40,000 and 150,000, I include in the sample cities or small groups of cities where the largest city outside of the market within a distance of 10 miles has a population below 10,000 and the population of the largest city within 20 miles does not exceed 25,000 people.[11] This selection rule serves to exclude candidate cities if they are part of a suburban sprawl or in a metropolitan area, which complicates the identification of market boundaries. Cities in tourist regions are also excluded since the resident population accounts for only a small share of the potential customer base. Neighboring cities are assigned to the same market if they lie within 10 miles of each other and either share boundaries with a candidate city or consist of Census tracts whose areas overlap with both cities.[12] As an additional check that the chosen markets are sufficiently geographically isolated from other cities in the region, I visually inspect each candidate market using regional maps. The resulting set of markets consists of 151 cities/groups of cities drawn from most U.S. states with a slight under-representation of the North East. Market size as measured by the included incorporated places' total population ranges from 41,352 to 142,303 people with an average market size of 74,367 people.

Further discretization is required to give meaning to the concept of a location within a market. First, the selected markets are divided into non-overlapping cells among which firms choose their optimal location. Following other recent applied work on the role of geography in microeconomic decision making (see, for example Davis 2001, Thomadsen 2003), I divide the markets into locations along Census delineated lines using Census tracts. While neighborhoods obviously change and shift over time, the use of Census tracts as cells comes closer to dividing the sample markets into coherent, internally homogeneous locations. Next, since consumers and firms are spread across the continuum of space of each of the Census tracts, instead of integrating over this geographic space, I place all

[11] All distances are computed as great circle distances according to the Haversine formula. Based on latitude-longitude coordinate data, the distance between two points, a and b, is given by:

$$d_{a,b} = 2R \arcsin\left[\min\left\{\left((\sin(0.5(lat_b - lat_a)))^2 + \cos(lat_a)\cos(lat_b)(\sin(0.5(lon_b - lon_a)))^2\right)^{0.5}, 1\right\}\right]$$

where R denotes the radius of the earth. See Sinnott (1984).

[12] Census tracts are small subdivisions of counties rather than cities, which have an average size of 4,000 people. The area of any given tract may therefore overlap with the area of more than one city. Census tracts that overlap with the sample cities are identified using the Census Bureau's geographic correspondence engine *MABLE/Geocorr*, available at `http://www.Census.gov/plue/`. This program includes a mapping utility that is capable of providing a comprehensive list of Census tracts whose area overlaps with each of the chosen markets. These overlapping tracts are included as part of a market unless the area of overlap contains an insignificant proportion of the tract's total population.

consumers and firms at the population-weighted centroid of their tract.[13] Each market is thus made up of a set of irregularly scattered point locations within the market's boundaries.

The classification of locations into product types is complicated by the irregularity in Census tract areas. Center-city neighborhoods are on average more densely populated compared to Census tracts at the outskirts of the city with more sparsely populated, larger areas. The differences in area between the small city center tracts and larger tracts at the outskirts imply that a firm's immediate competitors are not identified uniformly across locations if only competitors in the same location as the firm are included as such immediate competitors. Instead, I set the effect of competitors in the first distance band around a firm's location to be the same as that of competitors in the firm's own location. Given a short radius for the first distance band, this modification will not, in most cases, include tracts other than the tract in which the firm is located, but will only affect city centers where adjacent tracts are sufficiently close. Consequently, neighboring locations are defined to be all locations within a given distance range.

On average, a sample market consists of 21 tracts, ranging from markets with only eight tracts to markets with 49 tracts. The distance between tract centers within a market averages 3.5 miles using population-weighted centroids as tract centers. While the distance between a tract and its closest neighboring tract in the market is, on average, only 1.1 miles, the average distance to the furthest tract is 8.1 miles. Given the small distances that consumers are willing to travel to rent a video, these descriptive statistics indicate that the chosen markets are of an appropriate size to allow for spatial differentiation without being unrealistically large.

The use of Census tracts as market subdivisions also implies an inclusion of the city's surrounding population that resides in a tract at the edge of the city, but not within the city's official boundaries. This increases the average market size from 74,367 to 90,563 people. Given the geographic isolation imposed upon sample markets, the population living within the market represents a large fraction of the population residing in the general area and thus is a good approximation of the consumer base for which stores in that market compete.

[13] I obtain population-weighted centroids from the Census Bureau. They are used instead of the more standard area-weighted centroids to capture where the majority of the tract's population lives. For locations at the edge of a city, tracts tend to be large in area with an associated drop in population density and the area-weighted centroids generally lie at a greater distance from the remaining tracts' centroids than in the case of population-weighted centroids. The use of area-weighted centroids would therefore significantly overstate the attractiveness of these locations to firms in the form of greater distance from competitors in other locations in the city.

3.2 Video Rental Demand

Since store-level data on tape rentals are unavailable to me, I use the demographic characteristics of individual locations as a proxy for video rental demand. According to industry sources, total video demand is a function of the market's population, but varies as well across income levels, family status, and to a lesser extent age groups.[14] These demographic data are available from the Census Bureau's decennial Census of Population at a high level of geographic disaggregation, including Census tracts. The available firm-level data on location choices, which dates to 1999, is combined with demographic data from the 2000 Census of Population. In addition, a private data vendor, Advanced Geographic Solutions, provided data on tracts' business characteristics such as establishment counts across all industries and daytime working population. Comparable business summary statistics are generally not available from the Census Bureau at this level of geographic disaggregation. The overall business establishment counts are used in part to identify whether a tract is purely residential, namely if it does not contain any establishments. In such cases, the tract's population is included in market size indicators, but the tract itself is not included in the set of possible locations that firms can choose to enter.

Table 2 provides a summary of the key variables used to estimate the model. The demographic variables include the population in the store's chosen tract and in its immediate neighborhood, as well as the population residing in two bands around the chosen location. The use of the surrounding population reflects, in a reduced form, that people's shopping behavior is not confined to their immediate neighborhood, but may cover nearby areas. To capture income differences across locations, I use population-weighted average per-capita income of the tract and of locations around the tract, by distance band. The effects of other demand drivers, such as family status and proximity to a college or university, are more difficult to isolate at the tract level. Publicly available city planning records for a subset of the markets suggests that tracts with a high percentage of households with children or tracts that are home to a college tend to be protected by zoning ordinances, which prohibit firms from locating freely in such tracts. In the absence of detailed zoning data for the

[14] Hasting's Book, Music and Video, Inc. 1998 Annual Report states: "Key demographic criteria for Company superstores include community population, community and regional retail sales, personal and household disposable income levels, education levels, median age, and proximity of colleges or universities..." and the Video Software Dealers' Association (1998) claims: "The biggest demographic factor in determining a household's rental frequency ... is the presence of children. Almost three-fourths of all households with children rent at least once a month, while nearly a third rent at least once a week. Among households without children, 53% rent once a month or more and 21% rent once a week."

full set of markets, these residential tracts cannot be eliminated from the location choice set. Their inclusion in estimation confounds the role of family status and university locations as demand factors. The estimated effects of these demand drivers would not reflect the inherent attractiveness of such locations, but capture instead that regulation dictates for stores to never choose residential zones. Due to these difficulties in identifying potential locations, demographic characteristics other than population and per capita income are excluded from the estimation.

The attractiveness of a retail location stems partially from the easy accessibility and convenience that the location offers to consumers. While I do not have information on whether a store is located along a major commuting road or whether it is part of a strip mall receiving spillover business from other stores in the mall, a tract's business density is used as a proxy for its commercial character. The use of business density as a catchall proxy for the general business environment in the location also controls for the extent to which zoning laws enforce the residential nature of a tract. Location-specific costs to running a retail establishment take mainly the form of property costs and lease payments. Data on commercial rental costs is not available at as disaggregate a level as the Census tract, however; housing costs tracked by the Census Bureau are median residential rents. The use of median residential rents as a cost shifter in estimation had only limited success.[15] Consequently, the results laid out in section 4 use business density as the sole proxy for the role of the commercial environment in choosing a tract.

In summary, the demand shifters used in estimation include each location's population and per capita income, the population and average per capita income of each distance band around the location, and the location's business density.

3.3 Video Store Locations

Firm-level data on video store locations are obtained from American Business Disc 1999. This U.S.-wide semi-annual business directory contains information on establishment location, chain affiliation and lines of business and is derived from Yellow Page directories backed by phone inquiries.[16]

[15]The estimated effect of median rent levels on the likelihood of choosing a location was, as expected, negative, but insignificant.

[16]The information on firm counts and locations for the public video chains derived from the database is cross-checked against information contained in the respective firms' public SEC filings. For the six public chains in operation in 1999, the database contains more than 95% of the chains' outlets as per their 1998 fiscal year 10-K annual report. Furthermore, the total number of listed video retail establishments is 31,774. This number closely matches analyst

To match up store locations with Census tracts, each store's address is initially geo-coded. The resulting latitude-longitude coordinates are then assigned to the corresponding Census tract.

Firms' entry and location patterns vary significantly by market size and area. On average, 13.68 video stores compete in a market; the smallest has four stores and the largest 33 stores. 40% of the sample stores are affiliated with a chain, of which 42% belong to the two nationwide chains, Blockbuster Video and Hollywood Video. Due to the small size of the chosen markets, in 71% of the cases, chain stores operate only one store in any given market. In 38% of the markets in the sample, none of the firms operates more than one outlet, whereas in 36% of the markets, the maximum number of outlets under common ownership across all competitors is only two. On a per-market basis, the number of stores is therefore close to the number of firms. The market selection helps in justifying an assumption underlying the model, namely that each store makes an independent location decision based on idiosyncratic profitability draws. In larger markets, where chains operate multiple outlets in each market more frequently, it would be more appropriate to assume that each firm, as opposed to each outlet, receives an idiosyncratic profitability shock and chooses both the number and locations of outlets to operate.

At the tract level, both clustering in central locations as well as dispersion into locations at the city's edges can be observed. A significant fraction of the locations within a market is not chosen by any firm, but there are also locations that are selected by up to nine firms. As a result, some firms face many nearby competitors, the maximum number of firms that are located within half a mile of a firm's tract being ten stores. At the same time, some retailers choose isolated locations such that they do not face any competitors within a 10-mile radius. Within the market area, we observe that firms choose to locate both in the market's center and on its outskirts, the maximum distance from the city center reaching up to 15 miles.

The location statistics displayed in table 3 confirm the irregular distribution of firms. Figure 2 shows a map of one of the smaller sample markets, Great Falls, MT, chosen for this illustration due to its regular layout of Census tract neighborhoods. The city's boundaries overlap with 20 Census tracts, not all of which are depicted. The map shows clearly the variability in the tracts' areas and populations. Figure 2 also shows the number of competitors operating in each of the Census

estimates of the industry's size ranging from 30,000 to 35,000 outlets. See Advanstar Communications (various issues) and Video Software Dealers' Association (1998). Chain affiliation was complemented with firm listings contained in *Video Store Magazine*'s Top 100 Video Retailer surveys.

tracts as well as in three distance bands around one of the market's locations. The two concentric circles around the tract depict the bands containing immediately neighboring locations, as well as adjacent locations, which are between d_0 and d_1 miles away from the location.

For the estimation, firms are similarly placed into one of three distance bands and the distance cutoffs that define these bands are set to $d_0 = 0.5$ miles and $d_1 = 3$ miles, in accordance with Video Software Dealers' Association (1998) figures on customers' travel patterns. Thus, immediate competitors operate within one half of a mile from each other, neighboring competitors between one half and three miles from each other, and distant competitors include all remaining firms at a distance of more than three miles.[17]

The expected numbers of competitors in each of the distance bands are endogenously determined profit shifters. Based on distances between all $\mathcal{L}$ locations in the market, I define expected competitors as an $\mathcal{L} \times 3$ matrix where each entry represents the expected number of entrants into the market multiplied by the equilibrium probability of choosing any one of the tracts in the first, second, or third band around the location under consideration.

The data available for estimation thus consist of a scattered set of point locations within a market, the number of stores operating at those locations, the distance between locations used to construct competitor sets, as well as the locations' demographic characteristics.

4 Results

4.1 Parameter Estimates

Table 4 displays the estimated parameters under the two assumptions about the number of potential entrants, namely a potential entrant pool of 50 firms and a potential entrant pool of twice the actual entry observed in the market. The table also contains estimated marginal effects for the exogenous demographic variables. The marginal effects are computed by numerically differentiating the location-choice probabilities with respect to each demographic variable. The percent response in

[17] Since the markets vary significantly in in the maximum distance between tracts, the band covering the most distant locations imposes only a minimum cutoff on the distance between a pair of tracts and all firms that compete at a distance of more than three miles from one another are assumed to have the same incremental impact on profitability. Given the localized nature of video rental demand, this assumption appears justified for the chosen cutoff value. Experimenting with the cutoff between the neighboring and remaining categories had only small quantitative effects on the results.

probabilities to a one-percent increase in each exogenous variable above its observed average across all locations in the data set is computed on a location-by-location basis. The reported marginal effects represent the average response in probabilities across markets and locations.[18]

Most of the parameter estimates are of the anticipated sign and are precisely estimated due to the large within-market variation in the location characteristics. Across the two specifications, population has a large and positive effect on payoffs, but this effect decreases significantly with distance. A one-percent increase in the location's population implies, for example, approximately a three-percent increase in the likelihood of choosing this location. In contrast, a one-percent increase in the most distant population in the market area increases the likelihood of choosing the location by only one to two percent. Business density has a negative effect on firm profitability, while average per capita income has the expected positive effect on profitability, both in the location itself as well as in the remainder of the market. As with population, however, income levels in the chosen location and in immediately neighboring locations where the store's customers are likely to reside have practical significance for profitability. Per-capita income levels in the most distant locations have marginal effects on the likelihood of choosing a location of approximately 0.9%, about half of the marginal effect of average per-capita income levels in the chosen location and neighboring locations in the first distance band.

As expected, the presence of competitors has a negative effect on payoffs. This effect decreases significantly, however, with distance. For example, the presence of an additional competitor within half a mile from a firm's location has a payoff effect that is approximately 70% stronger than the effect of an additional competitor within one half to three miles who in turn would have a 52 to 66% stronger effect than a competitor located more than 3 miles away in the market. Thus, incentives for firms to differentiate are strong: spatial differentiation can effectively shield one's profit from a large number of the rivals operating in the same market.

The impact of changing the size of the potential entrant pool is most pronounced in the estimate of the mean of the market-level effect, μ. If only a small share of the potential entrants enters the market, the market-level effect ξ adjusts relative to the outside option's mean profitability of 0 to reflect the revealed low attractiveness of such a market. These realizations for ξ then yield a lower mean market-level effect. The results indicate a significantly lower estimate for μ in the

[18] Marginal effects cannot be computed for the endogenously determined expected number of competitors in the chosen location and in surrounding locations.

case of the 50-firm potential entrant pool relative to the pool size set to twice the actual number of entrants since with 50 potential entrants, in most markets the fraction of firms that decides to enter is smaller than in the alternative scenario.

The location-specific component of payoffs varies, in contrast to ξ, only with the actual number of entrants, $\mathcal{E}$, rather than $\mathcal{F}$. The parameters that determine these location-specific payoffs are affected only to the extent that the contribution of the stochastic term in payoffs changes with the number of potential entrants. As the size of the potential entrant pool rises, for example, a larger number of draws from the random payoff component ε is taken. With a larger set of ε-draws, but a fixed number of actual entrants equal to the observed firm count in each market, the most profitable entrants are likely to have higher unobserved profitability shocks than if the potential entrant pool were smaller. These selection effects may feed through to the parameter estimates of the observed location-specific payoff determinants. The results show, however, that the estimates of these parameters are quite similar across the two specifications, illustrating that such effects are small.

Figure 3 depicts the distribution of prediction errors based on the parameter estimates displayed in the fourth column of table 4. The mode of the distribution is slightly below zero. This skewed distribution is due to the fact that the Logit functional form assumption results in strictly positive probabilities for all location choices, even though many locations are ex-post not chosen by any firm. At the extreme, we observe some prediction errors that are rather large. On average, however, the included demographic characteristics and competitive effects predict location patterns fairly well.

The model places some strong assumptions on the underlying location-specific errors. In particular, it assumes that the unobserved profitability of a location is uncorrelated with that of neighboring locations. To examine the importance of spatial patterns in the unobservables, I conduct a test for spatial autocorrelation among the prediction errors that result from the model. I compute Moran's I statistic measuring the correlation between the prediction errors and a spatially weighted average of prediction errors in neighboring locations. The maximum value the test statistic takes under alternative specifications of the spatial weighting matrix is 0.062. There is thus little evidence of spatial correlation among the prediction errors. A full treatment of, and test for, spatial correlation of the unobserved profitability shocks, as opposed to choice prediction errors, requires the addition of a location-specific unobservable component to profits, which could be correlated with that of spatially close locations. Incorporating such location-specific unobservable attributes increases the computational burden of the model significantly and is left as an extension.

Figure 4 shows the empirical distribution of the market-level effects for a 50-firm potential entrant pool, as implied by the equilibrium condition that the predicted number of entrants equals the actual number of entrants in each of the markets. Figure 4 compares these standardized market-level effects to the assumed normal distribution for ξ. While the empirical distribution puts more weight on the center than the theoretical distribution, it approximates a bell curve.

4.2 Illustration of Results

The estimated parameters of the location-choice model suggest that at the level of the neighborhood, the competitive interaction between firms is strong. The lower competitive effects between more distant competitors indicates that firms exploit the geographic dispersion in their demand to lessen the competitive interaction with more distant rivals. Accordingly, we would expect that this lower degree of rivalry induces additional stores to enter as the market area and scope for differentiation grow. To quantify the importance of product characteristic choices in the entry process, I perform a counterfactual exercise that considers the role of the overall size of the characteristic space, here simply the geographic dispersion of demand, in affecting entry into the market.

The case of physical location in clearly delimited markets lends itself to conducting such investigations into the effect of individual features of the product-type space, such as its area and the distribution of consumers within the space. For other forms of product differentiation, the maximum degree to which firms can differentiate and how consumers are distributed across the various product type locations is not as easily observable. Other exercises of interest might include an investigation of the effect on entry of governmental regulation that restricts the extent of product differentiation between firms. In the context of spatial differentiation, such regulation most commonly takes the form of zoning ordinances that limit firms' abilities to locate freely within the entire area of a city. Similar examples from other contexts include licensing or minimum safety standards.

To isolate the effect of the extent of spatial dispersion in demand on market structure, one needs to recognize that as a city grows in size, not only does the city spread out spatially, but its population increases as well. Simply comparing predicted entry patterns across the sample markets that vary significantly in size thus does not allow us to separate the effect of the increased scope for spatial differentiation from the effect of the overall increase and scatter in population and thus market demand. To separate the contribution of each of these factors on the number of entrants that a market can support, I compare entry under two city growth scenarios.

The first scenario allows a city to grow in population only, holding its geographic layout fixed. In this case, firms' scope for spatial differentiation does not change since the total area of the city does not grow in proportion to the population. To do so, I take one of the smallest sample cities, Jamestown, NY, with twelve Census tract locations, and artificially increase Jamestown's population in increments of 1,500 people. The growth process leaves the number of locations, their layout, and the relative population shares across locations unchanged. As the population rises, it is thus only the population density in the twelve locations that increases, leaving the area that the city occupies unchanged.

Predicted entry under this city expansion path is then contrasted with entry that would occur were the city to grow both in population and area. While it is difficult to simulate how Jamestown, NY specifically would expand if it grew along both of these dimensions, the cross-section of sample markets can be used as a proxy for this growth path. The sample markets are suitable for this purpose since they span a range of market sizes from Jamestown, NY at the lower end with a population of 52,583 to larger markets such as Fort Collins, CO, with a population of 178,070. Furthermore, the larger cities in the sample naturally cover larger areas than the smaller cities and can thus serve to represent how Jamestown may look like were it to grow in population to their level. Based on the estimated parameters for the 50-firm potential entrant pool, I compute the expected number of entrants for the above city-growth scenarios. To do so, I integrate over the numerical distribution of the market-level effect ξ and find predicted location probabilities and entrants that are consistent with the market-level effect, the potential number of entrants into the market, and the market's exogenous characteristics. To abstract from cross-market and cross-location variations in business density and per-capita income that could drive entry patterns, I set these variables equal to the business density and per-capita income in Jamestown for all locations in the data.

Two opposing effects drive entry into a market when the spatial dimension of city growth is removed. The first effect comes from intensified competition. If the market area does not grow with a city's population, firms cannot spread out in space any further, decreasing the incentive for additional entry. The second, countervailing effect arises due to the fact that population becomes more dense within the given market area and firms will find a larger number of consumers in the immediate neighborhood of their store. The increased access to nearby consumers thus increases the incentive for additional entry into the growing city relative to a city that grows in both population and space. The net effect of these two forces determines whether entry into a city with a fixed market area exceeds or falls short of entry that we observe in markets that grow in population and area.

Figure 5 illustrates the role of geographic dispersion on expected entry as cities grow in population. In both panels, the scattered points correspond to predicted entry into the actual sample markets. The solid line represents the growth path of average predicted entry into the expanding Jamestown market, while the dotted lines denote the corresponding 95% confidence bands for entry.[19] The model predicts an approximately linear relationship between market size and equilibrium number of competitors for the specific range of video store market structures covered by the sample.

To separate the competition and demand effects of city growth, the top panel displays entry predictions assuming that the impact of population on payoffs does not vary by distance band. In particular, I set the three population parameters equal to the estimated parameter on population in the 0.5-to-3-mile distance band. As a city grows, the additional population then has the same impact on payoffs, regardless of where the population is located within the market. The chart demonstrates the effect of increasing a market's geographic space on firms' ability to capture localized market power by spatial differentiation. By the time Jamestown has grown in population to 150,000, allowing firms to also scatter in space amounts to an increase in the expected number of entrants of approximately ten stores. The difference between the two paths of expected entrants thus represents the contribution that the increased scope for spatial differentiation among firms makes to the number of firms that can profitably co-exist in growing markets.

Once one recognizes, however, that, as a city grows in space, customers at one end of the city are less likely to frequent a store at the other end of the city, the importance of the additional scope for differentiation decreases. The lower panel of figure 5 shows entry predictions that take both the population and competition effects into account. The estimated parameters for the entry and location choice model imply a localized pattern to the role of population in driving payoffs. The population in the immediate neighborhood of a firm's location has a higher payoff effect than the population in the remaining two, more distant, bands. The implication of this pattern on entry is that once the spatial aspect of city growth is removed, increased access to population

[19] The confidence bands are derived using bootstrap methods by predicting entry under 500 draws from the estimated parameter distribution for the 50-firm potential entrant pool.

in the immediate neighborhood increases payoffs, but this contribution falls short of the effect of increased competitive intensity on payoffs. As a result, predicted entry into the sample markets exceeds predicted entry into the growing Jamestown market with a fixed area, on average. The competition effect thus dominates. On net, however, allowing the area of the city to increase with its population does not lead to very significant increases in the predicted number of entrants; most of the predicted entry values for the actual sample markets fall within the 95% confidence band of predicted entry under the fixed market area. The results thus indicate that the size of the market area by itself has only limited implications for payoffs and consequently entry, probably since video retailing is an example of an industry where consumers' willingness to travel a long distance to a video store is low and demand is very local.

5 Conclusion

This paper has presented a framework for incorporating endogenous product-type choices into firms' entry decisions and has measured the subsequent impact on market structure for the video rental industry. The usefulness of the model lies in the fact that many applications of firm interaction involve a strong element of product differentiation. The results show evidence of strong incentives for spatial differentiation in the video rental industry, an industry where location is a major source of product differentiation among firms. Firms' abilities to capture localized market power are found to significantly increase with the size of the market space. As the market area grows, however, population spreads out as well limiting the benefits of such spatial differentiation. In this particular application where consumers exhibit strong preferences over the characteristics of the product they consume, I find that an expansion of the characteristics space and an associated increased dispersion of consumer preferences over characteristics induces little additional entry, suggesting that localized market power increases only slightly with an increased scope for product differentiation.

The model predicts a near-linear relationship between market size and the equilibrium number of competitors. In homogeneous product markets, this relationship suggests that market size and the number of competitors has grown to the point where oligopoly margins have disappeared and the threshold number of customers a store needs to enter profitably remains constant. This concept of an entry-threshold was coined by Bresnahan and Reiss (1991) as an indicator of the competitiveness of concentrated markets in the absence of price and quantity data to measure price elasticities. In a differentiated products setting, however, constant entry thresholds across markets of varying market structures may simply indicate that product differentiation offsets competitive decreases in margins with larger numbers of competitors. The significance of the location-choice model is that it permits an empirical analysis of the importance of such product differentiation in entry decisions at a disaggregate neighborhood level.

The modeling difficulties that arise due to the complexity of multiple agents' product differentiation decisions have been addressed by using an imperfect information framework in a static profit maximizing context. As in equilibrium models of social interactions,[20] firms endogenously sort into different locations depending on their expectation of the intensity of competition produced by rivals' simultaneous choices. The incomplete information framework entails some significant differences from its complete information counterpart, including an increased ease of deriving equilibrium location conjectures. A second distinguishing characteristic of the imperfect information framework is the possibility of ex-post regret: a firm's choice of location based on its assessment of its competitors' likely choices may ultimately not be the optimal choice once its competitors' actual choices are realized. The allowance for possible ex-post regret corresponds better to real-world environments and decision-making by firms. This holds in particular in cases where largely unobservable or immeasurable firm-specific capital contributes significantly to firm profitability.

Market structure is the equilibrium outcome of firms' simultaneous location choices based on single-period payoff comparisons. The motivation behind this albeit restrictive set-up is twofold. One reason is that long histories of firms' sequence of moves and market conditions are difficult to obtain. A second, more significant reason is that dynamic models of firm and industry evolution such as the one suggested by Ericson and Pakes (1995) are computationally complex and difficult to estimate for large markets with many product-types and many competitors. Similar to other empirical models of firm entry, the static model is assumed to approximate the repeated firm interaction that characterizes the evolution of an industry. Recent work on empirically implementing dynamic games of firm interaction may be able to evaluate the validity of this assumption.

[20] For a survey of recent developments in the specification and estimation of interactions-based models, see Brock and Durlauf (2001). Examples of empirical applications of endogenous sorting models into neighborhoods include Bayer, McMillan, and Rueben (2004) and Timmins (2003).

Appendix

The location-choice model results in a set of equilibrium location conjectures defined by equation 2.8, which map into firms' optimal strategies. Here, I briefly discuss existence and uniqueness properties of the equilibrium. The uniqueness of the equilibrium can be established analytically only for simple market layouts. I proceed by discussing numerical simulation evidence on the incidence of multiple equilibria under alternative parameter values for the competitive interaction effects.

Equation 2.8 sets up a continuous mapping from the $\mathcal{L}$-dimensional simplex into itself. Due to the constraint that probabilities sum to one, the system reduces to $(\mathcal{L}-1)$ equations in $(\mathcal{L}-1)$ unknown conditional location probabilities. Since firms' own conjectures are contained in the probability simplex and are continuous in competitors' expected behavior, the existence of at least one solution to the system of equations follows immediately from Brouwer's Fixed Point Theorem. To establish the uniqueness of such a solution to the system of conditional location choice probabilities,

$$\Psi(\boldsymbol{p}, \boldsymbol{X}, \mathcal{E}) = \boldsymbol{p} - F(\boldsymbol{p}, \boldsymbol{X}, \mathcal{E}) = 0, \tag{A.1}$$

it is sufficient to show that the matrix of partial derivatives of Ψ with respect to $\boldsymbol{p}$ is a positive dominant diagonal matrix, or that

1. $\dfrac{\partial \Psi_l}{\partial p_l} > 0$
2. $\left|\dfrac{\partial \Psi_l}{\partial p_l}\right| \geq \sum_{k \neq l} \left|\dfrac{\partial \Psi_l}{\partial p_k}\right|$

Consider first the simplest example of a 2×2 city that allows for spatially differentiated competition by letting the effect of competitors in a given location to be different from that of competitors in the remaining three locations.[21] Normalizing $p_4 = 1 - (p_1 + p_2 + p_3)$, the matrix of partial derivatives for this specific example contains the elements

$$\begin{aligned} \frac{\partial \Psi_l}{\partial p_l} &= 1 - (\mathcal{E} - 1)p_l(\gamma_0 - \gamma_1)(1 - p_l + p_4) \\ \frac{\partial \Psi_l}{\partial p_k} &= -(\mathcal{E} - 1)p_l(\gamma_0 - \gamma_1)(-p_k + p_4) \\ & \quad k \neq l,\ l, k = 1, 2, 3. \end{aligned} \tag{A.2}$$

Assuming without loss of generality that $p_4 = \min(\boldsymbol{p})$, conditions 1. and 2. hold in the case of more than one entrant provided that $\gamma_0 < \gamma_1$. Consequently, the location choice game for the 2×2 city has a unique equilibrium as long as immediate competitors that are located in the same cell drive profits down by more than more distant competitors. This simple example is suggestive of settings that entail multiple equilibrium location strategies. In particular, if competition between firms were to actually intensify as they move further away from each other, there may be many locations in which a firm would face the same expected number of distant competitors and thus

[21] Expected profits for this example are thus given by

$$\mathbb{E}[\Pi_{fl}] = \xi + \boldsymbol{X}_l\boldsymbol{\beta} + \gamma_0(1 + (\mathcal{E} - 1)p_l) + \gamma_1(\mathcal{E} - 1)\sum_{k \neq l} p_k + \varepsilon_{fl}$$

for $l, k = 1, ..., 4$.

an identical competitive environment. Similarly, uniqueness may break down in the case where there are positive externalities to clustering, that is γ_b is positive. Non-uniqueness arises in these scenarios in particular if there are only little or no differences in locations' demographic make-up to induce additional variation in payoffs across locations.

The profit function used in estimation has three distance bands that defined over larger sets of locations. For this more general payoff function, the elements of the matrix of partial derivatives are significantly more complex as they involve locations in additional distance bands. Allowing for three distance bands and a total of $\mathcal{L}$ locations, the partial derivatives of Ψ are given by

$$\begin{aligned} \frac{\partial \Psi_l}{\partial p_l} &= 1-(\mathcal{E}-1)p_l[(\mathbb{I}^1_{l\mathcal{L}}(\gamma_0-\gamma_1)+\mathbb{I}^2_{l\mathcal{L}}(\gamma_0-\gamma_2))(1-p_l+p_{\mathcal{L}}) \qquad \text{(A.3)} \\ &\quad +(\gamma_1-\gamma_2)(\sum_{k\neq l}(\mathbb{I}^2_{kl}(1-\mathbb{I}^2_{k\mathcal{L}})-\mathbb{I}^1_{kl}(1-\mathbb{I}^1_{k\mathcal{L}}))p_k)] \\ \frac{\partial \Psi_l}{\partial p_k} &= -(\mathcal{E}-1)p_l[(\mathbb{I}^1_{k\mathcal{L}}(\gamma_0-\gamma_1)+\mathbb{I}^2_{k\mathcal{L}}(\gamma_0-\gamma_2))(-p_k+p_{\mathcal{L}}) \qquad \text{(A.4)} \\ &\quad +(\gamma_1-\gamma_2)(\mathbb{I}^1_{kl}+\sum_{j\neq k}(\mathbb{I}^2_{jk}(1-\mathbb{I}^2_{j\mathcal{L}})-\mathbb{I}^1_{jk}(1-\mathbb{I}^1_{j\mathcal{L}}))p_j)] \\ &\quad k\neq l,\ l,k=1,...,\mathcal{L}-1, \end{aligned}$$

where $p_{\mathcal{L}}$ has been normalized to $1-\sum_{l\neq\mathcal{L}} p_l$ and, as before, $\mathbb{I}^b_{kl}=1$ if $d_b \leq d_{kl} < d_{b+1}$ and 0 otherwise. The partial derivatives are a function of the vector of location probabilities that depends on the dispersion of locations within the market in a complicated way. For the general payoff function, it is thus difficult to establish analytical conditions that guarantee that the matrix of partial derivatives has a positive dominant diagonal. The functional form of the partial derivatives suggests, however, that exogenous determinants of $\boldsymbol{p}$ and the layout of locations relative to each other in a given market are critical factors in the existence of a unique set of location probabilities.

To investigate the sensitivity of the equilibrium to within-market variation in exogenous demographic attributes, I apply the model to a simulated data set of markets that differ in the amount of variation in demographic attributes.[22] Equilibrium conjectures are found numerically using the method of successive approximations where the fixed point results from successively improving upon an initial guess for the probability vector until the probabilities solve the system of equations in 2.8. For a given market and set of parameter values, I compute the equilibrium for alternative starting values for firm conjectures. I consider an equilibrium to be unique if successive approximations to the equilibrium always converge to the same solution, independent of the initial starting values.

The simulations show that as long as γ_b is less negative for more distant bands (competitive interaction becomes weaker with distance), the fixed point algorithm converges to a single equilibrium, even if locations are fully homogeneous in exogenous attributes. In the case where γ_b becomes more negative with distance (competition intensifies the further competitors are from each other), the simulations converge to a single equilibrium only if there is variation in exogenous attributes or the number of locations is large. Otherwise, multiple equilibria arise.

[22] The simulations focus only on the case where γ_b is negative, that is where geographic proximity to other competitors decreases profits due to increased competition, rather than on the case of positive spillovers from geographic proximity to other firms.

Expressions A.3 and A.4 as well as the simulation evidence thus suggest that there are two major sources that lead to uniqueness in this model. First, heterogeneity in the demographic attributes of a firm's location and in the demographic attributes faced by neighboring and distant competitors of the firm, but not the firm itself, provide exogenous variation in profits that allows for a distinction of locations with similar sets of expected competitors. Second, the irregular dispersion of locations over the market area implies that the sets of locations that define immediate, neighboring, and distant competitors differ across locations. For markets with a large number of locations, there are few locations that have the same sets of competitors in the various distance bands. Furthermore, if competitive rivalry is strongest between competitors that are more rather than less alike in terms of geographic proximity, implying a γ_b that decreases with distance, the numerical simulations consistently result in a single equilibrium across all simulation runs, mirroring the analytic results for the case of a 2×2 city.

References

Advanstar Communications (various issues): *Video Store Magazine.* Santa Ana, CA.

Aguirregabiria, V., and P. Mira (2004): "Sequential Estimation of Dynamic Discrete Games," Mimeo, Boston University, Boston, MA.

Ahn, H., and C. Manski (1993): "Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations," *Journal of Econometrics*, 56, 291–321.

Andrews, D., S. Berry, and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Locations," Mimeo, Yale University, New Haven, CT.

Augereau, A., S. Greenstein, and M. Rysman (2004): "Coordination vs. Differentiation in the Adoption of High Speed Modems," Mimeo, Boston University, Boston, MA.

Bayer, P., R. McMillan, and K. Rueben (2004): "An Equilibrium Model of Sorting in an Urban Housing Market," Mimeo, Yale University, New Haven, CT.

Berry, S. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60, 889–917.

——— (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Equilibrium," *Econometrica*, 63, 841–890.

Berry, S., and J. Waldfogel (1999): "Free Entry and Social Inefficiency in Radio Broadcasting," *RAND Journal of Economics*, 30(3), 397–420.

Bresnahan, T., and P. Reiss (1988): "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531–553.

——— (1991): "Entry and Competition in Concentrated Markets," *Journal of Political Economy*, 99, 977–1009.

Brock, W., and S. Durlauf (2001): *Interactions-Based Models.* In: "Handbook of Econometrics," J. Heckman and E. Leamer (Eds.), Amsterdam: North-Holland.

Ciliberto, F., and E. Tamer (2004): "Market Structure and Multiple Equilibria in Airline Markets," Mimeo, North Carolina State University, Raleigh, NC.

Cotterill, R. W., and L. E. Haller (1992): "Barrier and Queue Effects: A Study of Leading U.S. Supermarket Chain Entry Patterns," *Journal of Industrial Economics*, 40, 427–440.

Davis, P. (2001): "Spatial Competition in Retail Markets: Movie Theaters," Mimeo, Sloan School of Business, MIT, Cambridge, MA.

Einav, L. (2003): "Estimating an Equilibrium Model of the Release Date Timing Game," Mimeo, Stanford University, Stanford, CA.

Ellickson, P. (2004): "Supermarkets as a Natural Oligopoly," Mimeo, Duke University, Durham, N.C.

Ericson, R., and A. Pakes (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62, 53–82.

Hasting's Book, Music and Video, Inc. (1998): *1998 Annual Report*.

Hotelling, H. (1929): "Stability in Competition," *Economic Journal*, 39, 41–57.

Manski, C. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60, 531–542.

Mazzeo, M. (2002): "Product Choice and Oligopoly Market Structure," *RAND Journal of Economics*, 33.

McKelvey, R. D., and T. R. Palfrey (1995): "Quantal Response Equilibria for Normal Form Games," *Games and Economic Behavior*, 10, 6–38.

Pakes, A., M. Ostrowsky, and S. Berry (2004): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," Mimeo, Harvard University, Cambridge, MA.

Pesendorfer, M., and P. Schmidt-Dengler (2003): "Identification and Estimation of Dynamic Games," Mimeo, London School of Economics, London, UK.

Reiss, P. (1996): "Empirical Models of Discrete Strategic Choices," *American Economic Review*, 86(2), 421–426.

Rust, J. (1994): *Estimation of dynamic structural models*. In: "Advances in Econometrics - Sixth World Congress," C. Sims (Ed.), Cambridge: Cambridge University Press.

Salop, S. (1979): "Monopolistic Competition with Outside Goods," *Bell Journal of Economics*, 10, 141–156.

Seim, K. (2001): "Spatial Differentiation and Market Structure: The Video Retail Industry," unpublished Ph.D. dissertation, Yale University.

Sinnott, R. W. (1984): "Virtues of the Haversine," *Sky and Telescope*, 68, 159.

Stavins, J. (1995): "Model Entry and Exit in a Differentiated-Product Industry: the Personal Computer Market," *Review of Economics and Statistics*, 76, 571–584.

Sutton, J. (1991): *Sunk Cost and Market Structure*. Cambridge: MIT Press.

——— (1998): *Technology and Market Structure*. Cambridge: MIT Press.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Models with Multiple Equilibria," *The Review of Economic Studies*, 70, 147–165.

Thomadsen, R. (2003): "The Effects of Ownership Structure on Prices in Geographically Differentiated Industries," Mimeo, Columbia Business School, New York, NY.

Timmins, C. (2003): "Estimating Spatial Differences in the Brazilian Cost-of-Living with Household Location Choices," Mimeo, Yale University, New Haven, CT.

Toivanen, O., and M. Waterson (2001a): "Empirical research on discrete choice game theory models of entry: an illustration," *European Economic Review*, 44, 985–992.

——— (2001b): "Market Structure and Entry: Where's the Beef?," *University of Warwick Working Paper*, No. 593, forthcoming RAND Journal of Economics.

Video Software Dealers' Association (1998): *An Annual Report on the Home Video Market 1998.* Encino, CA.

Watson, R. (2004): "Product Variety and Competition in the Retail Market for Eyeglasses," Mimeo, University of Texas at Austin, Austin, TX.

Tables and Figures

Table 1: Descriptive Statistics, Markets and Locations

	151 Sample Markets		
	Mean	Minimum	Maximum
Market level			
Population, market	74,367	41,352	142,303
Population, main city	59,428	40,495	140,949
Population, all tracts in market	92,563	41,614	193,322
Largest Incorporated Place within 10 mi	2,618	-	9,972
Largest Incorporated Place within 20 mi	7,916	-	24,725
Tract level			
Number of tracts	21.13	8	49
Number of store locations	18.72	7	44
Tract population	4,380	247	32,468
Area (sqmi)	10.10	0.10	181.50
Average distance (mi) to other locations in market	3.49	1.08	8.05

Notes:

The largest incorporated place within 10 and 20 miles is relative to the centroid of the market's main city. The distance between locations within a market is computed as the distance between the tracts' population-weighted centroids. Demographic data is as of 1999.

Table 2: Tract-level Demographic Characteristics

	Mean	Minimum	Maximum
Demographic characteristics			
Population	4,417	247	20,163
Population, within 0.5 mi of tract	4,952	247	23,676
Population, 0.5 - 3 mi of tract	42,281	0	145,499
Population, 3 - 10 mi of tract	54,817	0	169,271
Per capita income, within 0.5 mi of tract	17,807	3,484	60,347
Per capita income, 0.5 - 3 mi of tract	17,413	0	38,934
Per capita income, 3 - 10 mi of tract	19,417	0	38,452
Business characteristics			
Establishment density per square mile	177.86	0.15	5239.48

Notes:

The tract's total population is placed at the population-weighted centroid. Population within different distance bands to the tract under consideration is computed as the sum of the population in tracts for which the distance to the considered tract's centroid falls within the specified range. Demographic data is as of 1999.

Table 3: Store Location Patterns, Sample Markets

	Mean	Minimum	Maximum
Firms, market	13.68	4.00	33.00
Store clustering			
Firms, tract	0.73	0.00	9.00
Firms, within 0.5 mi of tract	0.80	0.00	10.00
Firms, within 0.5 - 3 mi of tract	6.12	0.00	27.00
Firms, within 3 - 10 mi of tract	7.94	0.00	33.00
Location patterns within city's area			
Distance to city center (mi)[1]	3.02	0.02	14.96

Notes:

All stores are placed at the tract's population-weighted centroid. Competitors within different distance bands to a firm's location are computed as the number of firms in tracts for which the distance to the firm's tract falls in the specified range.

[1] The city center is taken to be the population-weighted centroid of the market's main city.

Table 4: Parameter Estimates, Entry and Location Choice Model

	Potential Entrant Pool =			
	2 x Total Entrants		50 Firms	
Variable	Coefficient (Std. Error)	Marginal Effect	Coefficient (Std. Error)	Marginal Effect
Population$_0$ (000)	1.8191	0.0333	2.1258	0.0393
	(0.1534)		(0.1764)	
Population$_1$ (000)	1.3109	0.0236	1.7349	0.0314
	(0.1200)		(0.1498)	
Population$_2$ (000)	0.6070	0.0121	1.1348	0.0227
	(0.1192)		(0.1486)	
Business density	-0.8077	-0.0155	-0.8889	-0.0173
	(0.1458)		(0.1477)	
Avg. Per-Cap. Income$_0$ (0000)	0.9309	0.0180	1.0380	0.0204
	(0.1136)		(0.1233)	
Avg. Per-Cap. Income$_1$ (0000)	1.0081	0.0193	0.9188	0.0178
	(0.2081)		(0.2043)	
Avg. Per-Cap. Income$_2$ (0000)	0.4851	0.0092	0.4884	0.0094
	(0.2512)		(0.2601)	
γ_0	-3.4520		-3.3853	
	(0.3111)		(0.3266)	
γ_1	-1.0103		-1.0087	
	(0.0745)		(0.0923)	
γ_2	-0.3448		-0.4870	
	(0.0738)		(0.0934)	
σ	3.5829		4.6760	
	(0.3110)		(0.4316)	
μ	-2.8764		-7.0364	
	(1.3425)		(1.5801)	

Notes:
Results based on 1999 demographic and firm data. Subscript 0 denotes the immediately adjacent locations to the chosen tract, within 0.5 miles in distance; subscript 1 denotes tracts at 0.5 to 3 miles in distance from the chosen tract; and subscript 2 denotes tracts at more than 3 miles distance from the chosen tract. Tract-level business density is defined as the number of establishments (0000) per square mile. γ denotes competitive effects, and σ and μ the estimates of the parameters of the distribution of ξ.

Figure 1: Impact on Profits of Competitors' Locations: Illustration

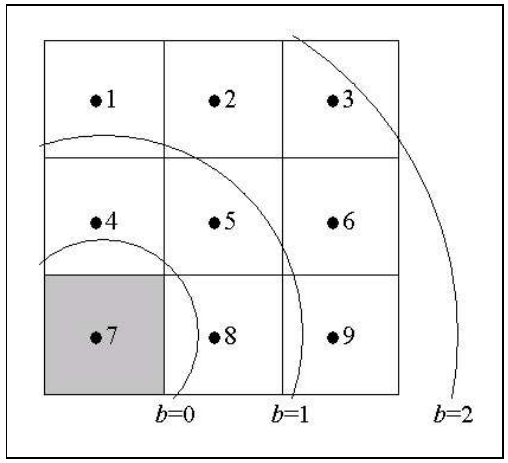


Figure 2: Sample Market - Great Falls, MT

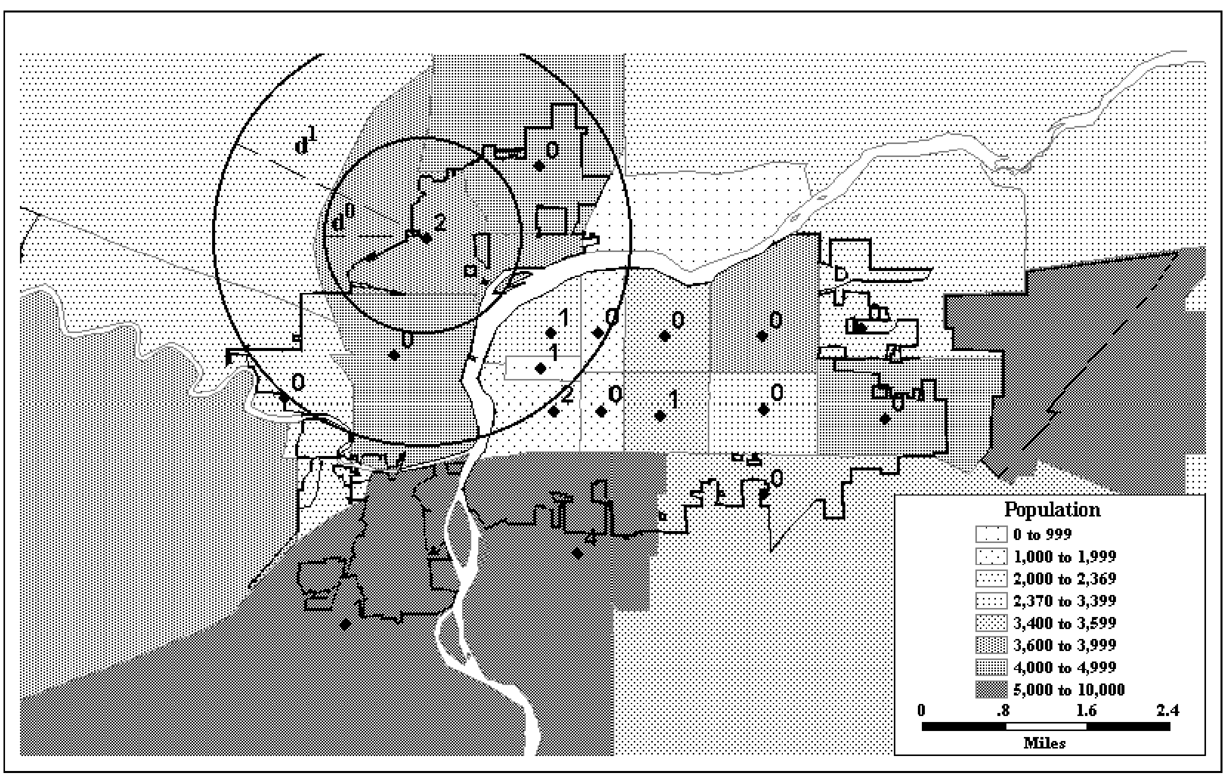


Figure 3: Distribution of Prediction Errors, Location Choice Probabilities

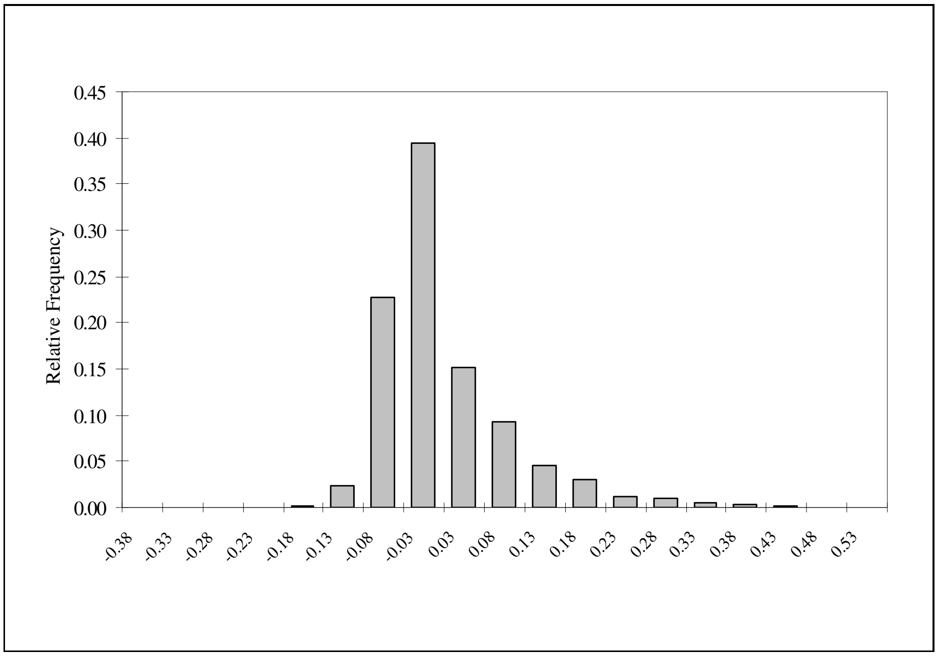


Figure 4: Distribution of Market-Level Effects

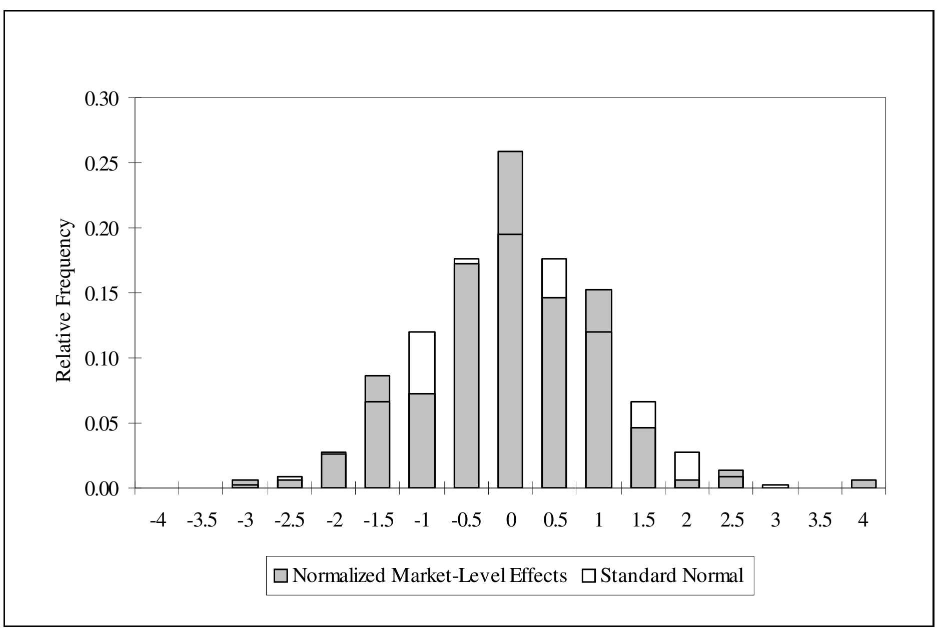
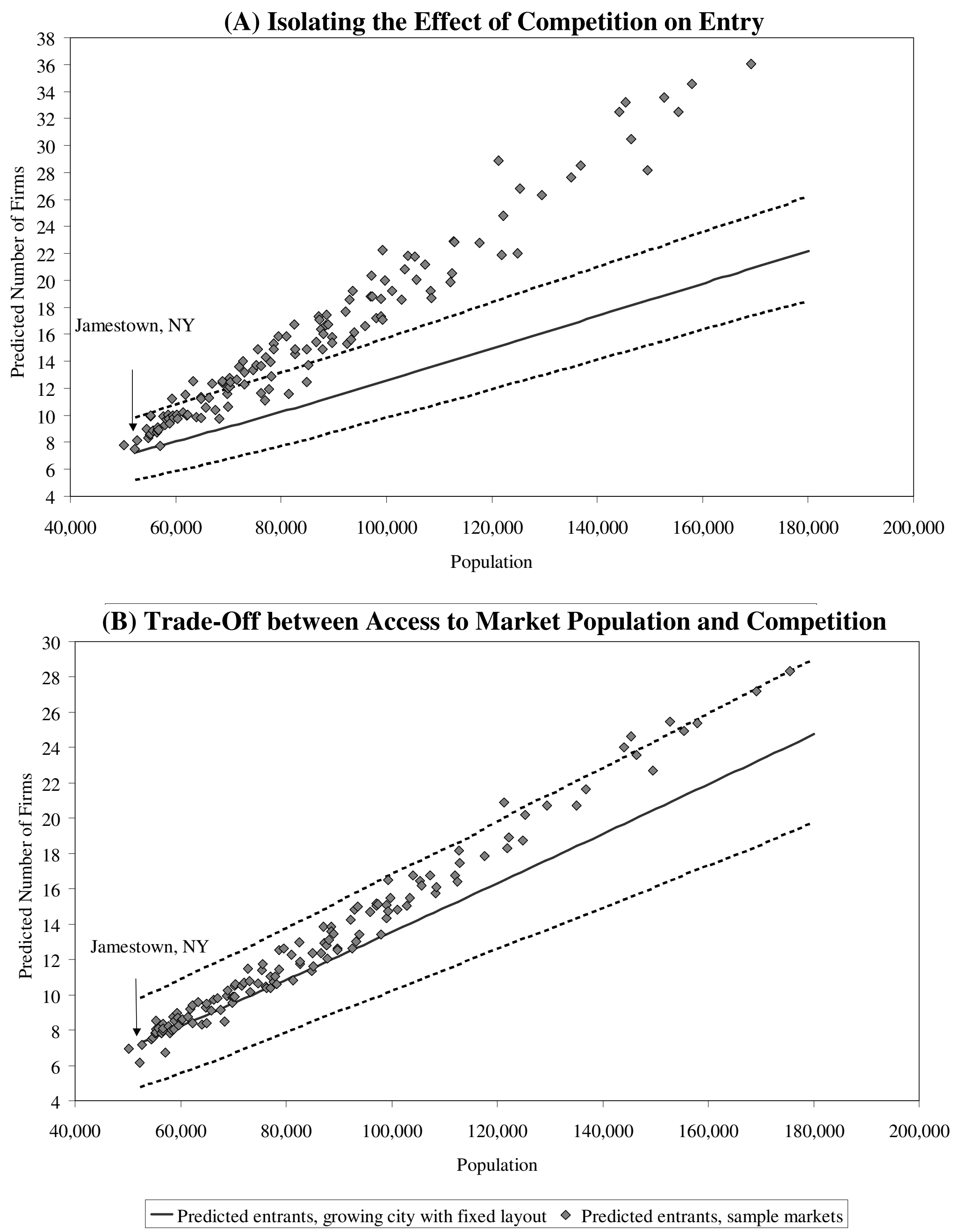


Figure 5: The Role of Spatial Dispersion on Entry



Note: Predicted entry patterns are based on estimated parameters for the 50-firm potential entrant pool. To isolate the effect of competition on entry in panel (A), all population parameters are set to the estimated effect on pay-off of population within 0.5 to 3 miles from the location.