

An empirical model of firm entry with endogenous product-type choices

Katja Seim*

I describe 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 nonuniform competitive effects on profits. I estimate the model for location choices in the video retail industry using a nested fixed-point algorithm solution. The results imply significant returns to product differentiation. Simulations illustrate the tradeoff between demand and intensified competition and the extent to which markets with more scope for differentiation support greater entry.

1. Introduction

■ Firms frequently compete in the characteristics of products that they offer to consumers. Economic theory going back to Hotelling (1929) and Lancaster (1966, 1979) has framed such product differentiation as competition between products located at different positions in an abstract characteristics space, with consumers having idiosyncratic preferences over these positions. A firm's location in characteristics space is a strategic variable, depending endogenously on the choices of its competitors.

Previous studies of entry have modelled the tradeoff between available demand and the intensity of competition faced by a new entrant, but the strategic importance of product positioning within a market has received less attention. Product differentiation allows a firm to better serve consumers' differing preferences and to acquire a degree of local market power. One determinant of the degree of market power is the scope for differentiation afforded by the product. For example, is local market power greater for automobile manufacturers with vast differentiation opportunities, or for makers of 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 insufficient demand for highly differentiated varieties?

I present an empirically tractable equilibrium model to analyze the determinants of firms' product positions. I focus on location decisions for a sample of video retailers. The empirical results support the intuition that firms use spatial differentiation to shield themselves from competition. Distant competitors affect payoffs significantly less than do nearby competitors. This competitive interaction helps to explain the location choices found in the data. The payoffs to

* University of Pennsylvania; kseim@wharton.upenn.edu.

This article 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, Rob Porter, Peter Reiss, John Rust, two anonymous referees, and the participants at various seminars for their thoughtful comments.

product differentiation motivate an analysis of the size of the characteristics space, measured by market area in my retail application, and its effect on market structure. I illustrate the effect of the geographic dispersion of demand on the entry process. As market area grows, firms can obtain more local market power, as there is more scope for spatial differentiation. However, payoffs from differentiation are lower as the population is more dispersed and demand falls. In the case of the video retail industry, which has highly localized demand, I find that the net effect of these two forces is that as population and market area grow, the number of firms increases only slightly. The extent of local market power generated by product differentiation depends on the interplay between differentiation possibilities and the strength of consumer preferences for particular product varieties.

My model addresses several difficulties in the empirical implementation of models of product differentiation and market structure. First, previous empirical work using discrete games to model product positioning decisions yields equilibrium strategies that are tractable only for a limited number of locations. Bresnahan and Reiss (1990, 1991) and Berry (1992) model firms' entry decisions as the equilibrium outcome of a discrete game played between potential entrants. Mazzeo (2002) extends this framework to predict firms' joint entry and quality-level choice.¹ An analytically appealing feature of his model is that firms have complete information about their rivals. A location distribution constitutes an equilibrium if no firm can increase profits by unilaterally changing its location. To confirm that a given configuration is an equilibrium entails comparing that configuration to every alternative configuration for every firm. Hence, computing an equilibrium configuration is difficult with large numbers of locations and firms. Mazzeo's (2002) application to the motel industry shows that, even with three quality levels, estimation is burdensome due to the large number of profit constraints that must hold.

In contrast to the complete-information framework, my model allows idiosyncratic sources of profitability, which are not observed by rivals. Examples include differences in cost or in intangible assets such as managerial talent, customer service, and inventory maintenance. Such factors affect profits, and they may not be observed by rivals who must separate the role of unobservables in a firm's success from other factors. I assume firms possess private information about their own profitability. Payoffs in a simultaneous-entry game thus depend on the firm's expectation of its rivals' location choices, as well as its idiosyncratic component. As Rust (1994) notes, the resulting Bayesian Nash equilibrium conjectures, which represent the probability of a particular strategy, can be derived more easily than in the complete-information framework, which solves for an exact equilibrium strategy. The incomplete-information model simplifies the computation of equilibrium strategies for a large-dimensional set of product types. Recent work on the estimation of discrete dynamic games (e.g., Aguirregabiria and Mira, 2006; Pakes, Ostrovsky, and Berry, 2005; and Pesendorfer and Schmidt-Dengler, 2003) also exploits information asymmetries to reduce the burden of computing dynamic entry and exit strategies.

My framework can accommodate a number of horizontal or vertical product differentiation choices, including discrete product attributes, combinations of attributes, or continuous attributes that can be discretized into distinct categories.² The model has been extended to a number of economic contexts. Watson (2005) uses a similar model to study firms' product variety choices in the context of eyeglass retailing. Augereau, Greenstein, and Rysman (forthcoming) analyze the importance of competition in a technology-adoption game between internet service providers and suggest differentiation between providers as an explanation for the absence of a single technological standard.

A second difficulty in empirical research on product-type choices stems from the need for detailed data on prices and quantities to estimate demand for various product space positions. Price and, in particular, quantity data are frequently not available. I do not have detailed outlet

¹ Related work includes Ellickson (2004), Stavins (1995), and Toivanen and Waterson (2005). For reviews of empirical research on discrete-choice game-theoretic models, see Toivanen and Waterson (2000) and Reiss (1996).

² For example, Mazzeo (2002) analyzes quality levels in the motel industry by discretizing a continuous range into three quality tiers.

pricing and rental data, so I exploit firms' location choices to make inferences about profitability. In doing so, I make the same type of assumption found in work by Sutton (1991, 1998), Bresnahan and Reiss (1990, 1991), and Berry (1992) that market structure reveals information about firms' underlying profits and the intensity of competition.

A last difficulty in estimating discrete strategic games is the possibility of multiple equilibria. Much of the literature (Berry, 1992; Bresnahan and Reiss, 1991; and Berry and Waldfoegel, 1999) circumvents multiple equilibria in individual players' actions by examining unique market and equilibrium outcomes.³ For example, entry models can predict a unique equilibrium number of entrants, but not the identity of the entrants. Asymmetric information in a discrete game does not necessarily guarantee a unique equilibrium. I show, however, that for a simplified version of my model, a unique equilibrium exists under reasonable restrictions on the model parameters. I present simulation evidence for the more general model.

The remainder of the article proceeds as follows. In Section 2, I present the model of entry and location choice and discusses econometric estimation. Section 3 describes the dataset. Section 4 presents the estimation results and counterfactual analyses of the effect of spatial differentiation on entry. In Section 5, I conclude.

2. Model

■ **Setup and payoffs.** I examine a firm's choice of product position among a set of discrete locations in characteristics space. To ensure that equilibrium strategies can be computed for many locations, my model is static. Accordingly, each firm makes its 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, firms are attracted to locations with favorable demand characteristics. Firm behavior also depends endogenously on the choices of other firms because increased competition at a given location adversely affects profits. Profits vary between firms in the same location due to differences in costs and other firm-specific factors. I assume that such differences are observed only by the firm itself. Entry and location choices are thus determined by the demand characteristics in each market location, firms' expectations of their competitors' 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 $\ell = 0, 1, \dots, \mathcal{L}^m$, where the decision not to enter is denoted by $\ell = 0$. Firm f 's location decision, for $f = 1, \dots, \mathcal{F}$, is denoted by \mathbf{d}_f , where $d_{f\ell} = 1$ if location ℓ is chosen and 0 otherwise.

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

$$\Pi_{f\ell}^m = \mathbf{X}_\ell^m \beta + \xi^m + h(\Gamma_{\cdot\ell}^m, \mathbf{n}^m) + \varepsilon_{f\ell}^m. \quad (1)$$

The first two terms represent demand characteristics that affect payoffs in location ℓ in market m . \mathbf{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

³ Tamer (2003), Andrews, Berry, and Jia (2004), and Ciliberto and Tamer (2006) instead derive upper and lower bounds for the probability of each nonunique outcome. Their estimation identifies bounds on the parameters of interest, rather than point estimates. Ciliberto and Tamer apply these procedures to airline route selection.

⁴ The model thus ignores one dimension of the decision-making process for multiproduct firms, namely the cannibalization of revenues of existing products when introducing a new product. The significance of this assumption depends upon the setting. I discuss the importance of chain stores in the sample markets in Section 3.

⁵ This specification is chosen to overcome the unavailability of firm-level market shares. Berry (1992) and Bresnahan and Reiss (1991) use similar payoff functions, interpreted as the product of a market-size term and a variable-profit term that depends on the number of competitors. One study that uses market share information is Berry and Waldfoegel (1999).

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. The next term, $h(\Gamma_\ell^m, \mathbf{n}^m)$, captures the effect on profits due to competition. Nonuniform 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 ℓ th column of Γ , Γ_ℓ^m represents the competitive intensity between rivals 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. In the case of spatial differentiation, the appropriate metric to measure the similarity between firms' products is physical distance. The impact on payoffs due to competition from other firms is thus a function of Γ_ℓ^m and \mathbf{n}^m , where \mathbf{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 are normalized to zero across firms and markets.

$\varepsilon_{f\ell}^m$ represents the idiosyncratic component of firm f 's profits from operating in location ℓ . As in Rust (1994), 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, but whose realization is private information. Players' information sets and types are defined by the following assumption.

Assumption 1 (Independent symmetric private values). Players' profitability types $\varepsilon_1^m, \dots, \varepsilon_F^m$ are private information to the players and are independently and identically distributed draws from the 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 in the previous literature. Profits depend only on the number of entrants at every location, and not on the entrants' identities.⁶ Symmetry implies that any two firms 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 two 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.

Assumption 2. $h(\Gamma_\ell^m; \mathbf{n}^m) = \sum_{k=1}^{\mathcal{L}^m} \gamma_{k\ell} n_k^m$.

Assumption 3. $\gamma_{k\ell} = \gamma_{k'\ell} = \gamma_b$ if $\mathcal{D}_b \leq d_{k\ell}^m, d_{k'\ell}^m < \mathcal{D}_{b+1}$, where \mathcal{D}_b and \mathcal{D}_{b+1} denote cutoffs that define a distance band.

Assumption 2 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 3 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 3 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 profits of firms in location ℓ , across locations in a market, and across markets.⁷ As a result, rivals located in tracts within the same distance band around location ℓ exert the same competitive pressure.

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

⁶ I maintain symmetry for computational reasons. One can 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 A potential problem with this approach is that the competitive pressure exerted by two firms located in a single location to the north of ℓ is the same as the competitive pressure of two firms, of which one operates in the cell north of ℓ and one in the cell south of ℓ . Typically, the first scenario would be more attractive to firm f than the second.

$$\Pi_{f\ell} = \xi + \mathbf{X}_\ell \beta + \sum_b \gamma_b N_{b\ell} + \varepsilon_{f\ell}, \quad (2)$$

where market superscripts m have been omitted to simplify the exposition. In (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 \mathcal{D}_1 , γ_1 the competitive effect of firms at a distance between \mathcal{D}_1 and \mathcal{D}_2 , and so forth. The total number of firms in distance band b is given by $N_{b\ell}$. Specifically, $N_{b\ell} = \sum_k \mathbb{I}_{k\ell}^b n_k$, where $\mathbb{I}_{k\ell}^b = 1$ if $\mathcal{D}_b \leq d_{k\ell} < \mathcal{D}_{b+1}$ and 0 otherwise. Summing $N_{b\ell}$ across distance bands yields the total number of competitors in the market, \mathcal{E} .

□ **Conjectures and equilibrium.** Due to imperfect information about its rivals' profitability, a firm can only form an expectation of their optimal location choices. Based on the expected competitor distribution across market locations, each firm will choose the location that maximizes its payoffs given its own type. The expected profit in location ℓ is

$$\begin{aligned} \mathbb{E}[\Pi_{f\ell}] &= \xi + \mathbf{X}_\ell \beta + \sum_b \gamma_b \mathbb{E}[N_{b\ell}] + \varepsilon_{f\ell} \\ &= \mathbb{E}[\bar{\Pi}_{f\ell}] + \varepsilon_{f\ell}, \end{aligned} \quad (3)$$

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

Assumption 2 that the number of rivals enters profits linearly simplifies the computation of expected competition. In contrast, Berry (1992) and Bresnahan and Reiss (1991) employ more flexible functional forms for $h(\cdot)$ that decrease in \mathbf{n} at a declining rather than constant rate. Relaxing Assumption 2 to incorporate a more flexible functional form for $h(\gamma, n)$ is a computational rather than conceptual difficulty, as it would involve more complicated numerical integration techniques.

I begin by analyzing location strategies for a given number of entrants. The determination of the total number of entrants is discussed at the end of this section. Due to the symmetry of rivals' types, firm f 's perception of firm g 's location strategy is the same for all competitors. The probability that competitor g chooses location ℓ , $p_{g\ell}$, is given by

$$p_{g\ell}(d_{g\ell} = 1 \mid \xi, \mathbf{X}, \mathcal{E}, \theta_1) = \Pr(\mathbb{E}[\bar{\Pi}_{g\ell}(\cdot)] + \varepsilon_{g\ell} \geq \mathbb{E}[\bar{\Pi}_{gk}(\cdot)] + \varepsilon_{gk}, \quad \forall k \neq \ell, \forall g \neq f), \quad (4)$$

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

$$\mathbb{E}[N_{b\ell}] = \sum_k \mathbb{I}_{k\ell}^b \mathbb{E}[n_k] = \sum_k \mathbb{I}_{k\ell}^b (\mathcal{E} - 1)p_{gk} + \mathbb{I}_{b=0}. \quad (5)$$

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 .

I assume that players' types, ε , are i.i.d. draws from a type I extreme-value distribution. 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. This implies that profitability is uncorrelated across firms within a given location, as well as across locations for a given firm. Thus the specification does not exhibit spatial correlation.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, profitability draws across locations are perfectly correlated and concentrated at the mean. In this case, there is no uncertainty about rivals' profitability, and McKelvey and Palfrey (1995) show that the outcome

⁸ To allow for more realistic substitution patterns across locations, random parameter coefficients or a more flexible error distribution could be used, which would raise significantly the computational complexity of finding equilibrium probabilities.

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 normalized to one. This results in multinomial Logit probabilities for firms' beliefs, conditional on the entry of \mathcal{E} firms, given by

$$p_{g\ell} = \frac{\exp(\mathbb{E}[\bar{\Pi}_{g\ell}])}{\sum_{k=1}^{\mathcal{L}} \exp(\mathbb{E}[\bar{\Pi}_{gk}])} \tag{6}$$

The equilibrium is a symmetric Bayesian Nash equilibrium describing the optimal response that maximizes the entering firm's expected payoff, 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 $\mathbf{p}_g = \mathbf{p}_f = \mathbf{p}^*$. A firm's vector of equilibrium conjectures over all locations ℓ is then defined by the following set of \mathcal{L} equations:

$$\begin{aligned} p_{\ell}^* &= \frac{\exp(\bar{\Pi}_{\ell}(\mathbf{X}, \mathbf{p}^*, \mathcal{E}, \theta_1))}{\sum_{k=1}^{\mathcal{L}} \exp(\bar{\Pi}_k(\mathbf{X}, \mathbf{p}^*, \mathcal{E}, \theta_1))} \\ &= \frac{\exp(\mathbf{X}_{\ell}\beta + \gamma_0 + (\mathcal{E} - 1) \sum_b \gamma_b \sum_j \mathbb{I}_{j\ell}^b p_j^*)}{\sum_{k=1}^{\mathcal{L}} \exp(\mathbf{X}_k\beta + \gamma_0 + (\mathcal{E} - 1) \sum_b \gamma_b \sum_j \mathbb{I}_{jk}^b p_j^*)} \quad \forall \ell = 1, \dots, \mathcal{L}, \end{aligned} \tag{7}$$

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 rivals' strategies into its rivals' 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.

As an illustration, Figure 1 depicts a square-shaped market with nine locations, grouped into three distance bands. Consider the payoffs to locating in cell 7. Immediate competitors are those in that location. The neighboring competitors in band 1 are located in cells 4, 5, and 8, while the most distant competitors are located in cells 1, 2, 3, 6, and 9. Given equation (4), $\mathbb{E}[\bar{\Pi}_7]$ equals

$$\mathbb{E}[\bar{\Pi}_7] = \xi + \mathbf{X}_7\beta + \gamma_0 + (\mathcal{E} - 1)(\gamma_0 p_7^* + \gamma_1(p_4^* + p_5^* + p_8^*) + \gamma_2(p_1^* + p_2^* + p_3^* + p_6^* + p_9^*)). \tag{8}$$

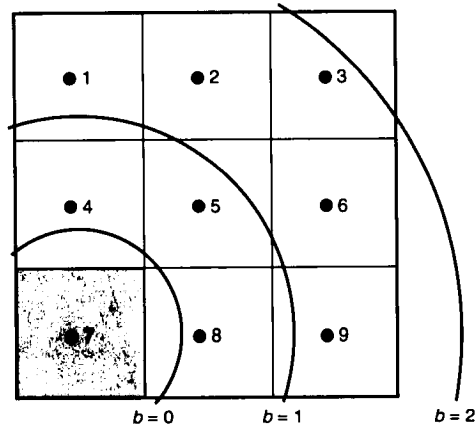
Assuming that the competitive impact of neighboring firms, γ_1 , exceeds that of more distant firms, γ_2 , the appeal of cell 7 lies primarily in its placement at the edge of the city with a small set of immediately adjacent locations and competitors. In contrast, a firm located in cell 5 will have many 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 5 will be more attractive than cell 7 if it grants easy access to most of the consumers living in neighboring locations. The equilibrium firm location pattern is determined by this tradeoff between demand and competitive pressures.

In equation (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 nonnegative 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

$$\text{Pr}(\text{entry}) = \frac{\exp(\xi) \left[\sum_{\ell=1}^{\mathcal{L}} \exp(\bar{\Pi}_{\ell}(\mathbf{X}, \mathbf{p}^*, \mathcal{E}, \theta_1)) \right]}{1 + \exp(\xi) \left[\sum_{\ell=1}^{\mathcal{L}} \exp(\bar{\Pi}_{\ell}(\mathbf{X}, \mathbf{p}^*, \mathcal{E}, \theta_1)) \right]} \tag{9}$$

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 (4) by ξ . The probability of entry is identical across competitors,

FIGURE 1
IMPACT ON PROFITS OF COMPETITORS' LOCATIONS: ILLUSTRATION



and as a result, the expected number of entrants is simply

$$\mathcal{E} = \mathcal{F} \cdot \Pr(\text{entry}). \quad (10)$$

Through (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 datasets include information on actual entrants, but not on those who consider entering but choose not to. Cotterill and Haller (1992) who study entry into grocery retailing, deal with this problem by setting the potential entrant pool equal to the number of major chains. Nonchain-affiliated retailers have a significant presence in the current application, and 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% 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 of Cotterill and Haller. I use a potential entrant pool of 50 firms, which for most markets is more than twice the number of actual entrants. I can compare the estimated parameter values under the two different assumptions on potential entrants.

To complete the game-theoretic model, the system of equations in (7) can be augmented by (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, I use the method of successive approximations to find it from an initial guess of the equilibrium.

□ **Estimation.** The augmented system of equations (7) and (10) is highly nonlinear, which is numerically difficult to solve, in particular if the initial guess of the number of entrants into a market is far from the ultimate solution to (10). Instead, in the empirical implementation, I simplify the entry decision by assuming that the expected number of entrants predicted by the model in (10) exactly 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 (10) equals the observed number of entrants. Equations (9) and (10) can be solved for ξ as a function of location characteristics, equilibrium location conjectures, and the number of potential and actual entrants:

$$\xi = \ln(\mathcal{E}) - \ln(\mathcal{F} - \mathcal{E}) - \ln \left(\sum_{\ell=1}^{\mathcal{L}} \exp(\bar{\Pi}_{\ell}(\mathbf{X}, \mathbf{p}^*, \mathcal{E}, \theta_1)) \right). \quad (11)$$

Upon substitution of the observed number of entrants in a market for \mathcal{E} , (11) yields a market-

specific realization for ξ . For this particular realization of ξ , the predicted number of entrants coincides with the observed number in each market. This approach of using an unobservable effect to induce an equivalence between actual and predicted numbers of entrants parallels the approach used by Berry (1994) and Berry, Levinson, 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 (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

$$L(\theta_1, \theta_2) = \prod_{m=1}^M \mathbf{p}_{\theta_1}(\mathbf{d}^m | \xi^m, \mathbf{X}^m, \mathcal{E}^m) g_{\theta_2}(\xi^m | \mathbf{X}^m, \mathcal{E}^m, \mathcal{F}^m), \quad (12)$$

where $\mathbf{d}^m = (d_1^m, d_2^m, \dots, d_{\mathcal{F}}^m)$ 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 (7) 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, \mathbf{p}^m .⁹ The equilibrium location-choice probabilities, together with \mathcal{E} and \mathcal{F} , feed into (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 estimate the parameters to best explain the observed firm behavior. The parameters include β 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 (12) using a Nelder-Mead 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 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 uniqueness 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, the model has a unique equilibrium, provided competitive intensity between firms decreases with distance. For a 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 intensity of competition decreases with distance or when locations differ in their characteristics, generating exogenous variation for competitors' expected location choices. There is variation in the sample locations' exogenous characteristics,

⁹ Explicitly solving for the equilibrium may be computationally burdensome for more complex models. Alternative approaches rely on initial nonparametric estimates of the respective expectations and equilibrium choice probabilities to derive players' optimal decisions and draw inferences about the underlying preferences. See Ahn and Manski (1993) and Aguirregabiria and Mira (2006).

as discussed in Section 3. Consequently, I do not impose any restrictions on γ in estimation, and I verify *ex post* that the equilibrium associated with the optimal parameter values is unique.

3. Data

■ I apply my model 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 rental prices 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, rental contract terms concerning the rental period, and drop-off convenience. The main avenue of differentiation arises, however, from spatial location, since the small price differences across stores make customers unwilling to travel a long distance. The spatial dimension of product differentiation is my main focus.¹⁰

□ **Sample markets.** Spatial differentiation will play a significant role in market structure only if the population, or available demand, is sufficiently large and geographically spread out for firms to 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 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 excludes candidate cities if they are part of a suburban sprawl or in a metropolitan area. I further exclude cities in tourist regions, since the resident population accounts for 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 neighborhoods whose areas overlap with both cities. As an additional check that the chosen markets are sufficiently geographically isolated, I 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 underrepresentation of the Northeast. As shown in Table 1, 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 nonoverlapping cells. Following recent applied work on the role of geography (such as Davis (forthcoming) and Thomadsen (2005)), I divide the markets into locations using Census tracts. Census tracts have an average size of 4,000 people. While neighborhoods change over time, Census tracts are cells that divide the sample markets into coherent, internally homogeneous locations.¹² Next, I place all consumers and firms at the population-weighted centroid of their Census tract.¹³ I prefer these to the more standard area-weighted centroids to capture where the majority of the tract's population lives. For locations at

¹⁰ Seim (2001) contains an extension of the model that allows firms to further differentiate by incorporating two types of firms, larger stores with a deeper inventory (proxied by whether stores are affiliated with a chain) and smaller stores.

¹¹ All distances are computed as great circle distances according to the Haversine formula. See Sinnott (1984).

¹² I identify tracts that overlap with the sample cities using the Census Bureau's geographic correspondence engine *MABLE/Geocorr*, available at <http://oseda.missouri.edu/plue/>. All overlapping tracts are included as part of a market unless the area of overlap contains an insignificant proportion of the tract's total population.

¹³ Population-weighted centroids are available from *MABLE/Geocorr*.

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 miles	2,618	—	9,972
Largest incorporated place within 20 miles	7,916	—	24,725
Tract Level			
Number of tracts	21.13	8.00	49.00
Number of store locations	18.72	7.00	44.00
Tract population	4,380	247.00	32,468
Area (square miles)	10.10	.10	181.50
Average distance (miles) to other locations in market	3.49	1.08	8.05

Note: 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 are as of 1999.

the edge of a city, tracts tend to be large 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. Area-weighted centroids would therefore overstate the attractiveness of these locations, in the form of greater distance from competitors in other locations. Each market thus consists of a set of irregularly scattered point locations within the market's boundaries.

The classification of locations into product types is complicated by irregularity in Census tract areas. Center-city neighborhoods are on average more densely populated and smaller than Census tracts at the outskirts of the city. The differences between the small city center tracts and larger tracts at the outskirts imply that a firm's immediate competitors would not be identified uniformly across locations if I included only those competitors in the same location as the firm. Instead, I set the effect of competitors in the first distance band around a firm's location to be the same as that of rivals in the firm's own location. Given a short radius for the first distance band, in most cases, this modification will not include tracts other than the tract in which the firm is located, but will only affect city centers where adjacent tracts are 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 8 tracts to markets with 49 tracts. The distance between tract centers within a market averages 3.5 miles. While the distance between a tract and its closest neighboring tract 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 statistics indicate that the chosen markets are appropriate to allow for spatial differentiation, without being unrealistically large.

The use of Census tracts as market subdivisions also includes 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 92,563 people. Given the geographic isolation imposed upon sample markets, the population living within the market is a large fraction of the population residing in the general area, and thus is a good approximation of the consumer base for which stores compete.

□ **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

TABLE 2 Tract-Level Demographic Characteristics

	Mean	Minimum	Maximum
Demographic Characteristics			
Population	4,417	247	20,163
Population, within .5 miles of tract	4,952	247	23,676
Population, .5–3 miles of tract	42,281	0	145,499
Population, 3–10 miles of tract	54,817	0	169,271
Per capita income, within .5 miles of tract	17,807	3,484	60,347
Per capita income, .5–3 miles of tract	17,413	0	38,934
Per capita income, 3–10 miles of tract	19,417	0	38,452
Business Characteristics			
Establishment density per square mile	177.86	.15	5,239.48

Note: 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 are as of 1999.

it 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 for individual 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 data 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 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 locations that firms can enter.

Table 2 provides a summary of the key variables used to estimate the model. The demographic variables include the tract population, as well as the population residing in two bands around the chosen location. The surrounding population reflects that people's shopping behavior is not confined to their immediate neighborhood, but may cover nearby areas. I also 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 suggest that tracts with many 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, 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 just reflect the inherent attractiveness of such locations, but also local regulation. Due to these difficulties, demographic characteristics other than population and per capita income are excluded from the estimation.

The attractiveness of a retail location stems partially from its accessibility and convenience to consumers. 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 catch-all proxy for the general business environment also controls for the extent to which

¹⁴ The 1998 annual report of Hasting's Book, Music and Video, Inc. states that "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."

zoning laws enforce the residential nature of a tract. Location-specific costs of running a retail establishment mainly take the form of property costs and lease payments. Data on commercial rents is not available at the Census tract level; however, housing costs tracked by the Census Bureau are median residential rents. The use of median residential rent as a cost shifter in estimation had only limited success.¹⁵ Consequently, the results described in Section 4 use business density as the only proxy for the commercial environment.

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.

□ **Video store locations.** Firm-level data on video store locations are obtained from *American Business Disc 1999*. This national semi-annual business directory contains information on establishment location, chain affiliation, and lines of business, and it is derived from Yellow Page directories backed by phone inquiries. I cross-checked the information on firm counts and locations for the public video chains derived from the database 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 estimates of the industry's size from Advanstar Communications' *Video Store Magazine* (various issues) and Video Software Dealers' Association (1998) that range from 30,000 to 35,000 outlets. To match up store locations with Census tracts, each store's address is 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. Forty percent of the sample stores are affiliated with a chain, of which 42% belong to the two nationwide chains, Blockbuster Video and Hollywood Video. Because the chosen markets are small, in 71% of the markets, chain stores operate only one store. 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 justify the assumption that each store makes an independent location decision based on idiosyncratic profitability draws. In larger markets, where chains operate multiple outlets, 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 the locations of outlets.

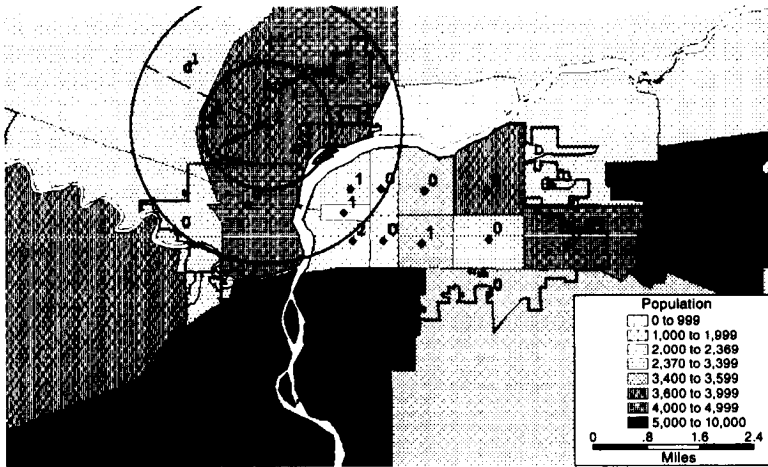
At the tract level, both clustering in central locations as well as dispersion into locations at the city's edges can be observed. A significant fraction of the locations within a market do not witness entry, but there are also locations that are selected by as many as nine firms. As a result, some firms face many nearby competitors. The maximum number of firms located within half a mile of a firm's tract is ten. At the same time, some isolated locations do not have any competitors within a ten-mile radius. Within the market area, firms locate both in the market's center and on its outskirts, the maximum distance from the city center reaching fifteen 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, Montana—chosen 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 clearly shows the variability in the tracts' areas and populations. Figure 2 shows the number of stores in each Census tract in Great Falls, as well as in three distance bands around one location. The two concentric circles around the tract depict the bands containing immediately neighboring locations, as well as adjacent locations, which are between \mathcal{D}_0 and \mathcal{D}_1 miles away from the location.

For the estimation, firms are placed into one of three distance bands. The distance cutoffs that define these bands are $\mathcal{D}_0 = .5$ miles and $\mathcal{D}_1 = 3$ miles, in accordance with Video Software Dealers'

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

FIGURE 2
SAMPLE MARKET: GREAT FALLS, MONTANA



Association (1998) figures on customers' travel patterns. Thus, immediate competitors are within a half-mile apart, neighboring rivals between a half-mile and three miles apart, and distant competitors include all firms more than three miles apart. Since the markets vary significantly 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. 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 local nature of video rental demand, this assumption appears justified. Experimenting with the cutoff between the neighboring and remaining categories had only small quantitative effects on the results.

The expected numbers of competitors in each distance band are endogenously determined. 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, as well as the locations' demographic characteristics.

TABLE 3 Store Location Patterns, Sample Markets

	Mean	Minimum	Maximum
Firms, market	13.68	4.00	33.00
Store Clustering			
Firms, tract	.73	.00	9.00
Firms, within .5 miles of tract	.80	.00	10.00
Firms, within .5-3 miles of tract	6.12	.00	27.00
Firms, within 3-10 miles of tract	7.94	.00	33.00
Location Patterns within City's Area			
Distance to city center (miles) ^a	3.02	.02	14.96

Note: 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.

^a The city center is taken to be the population-weighted centroid of the market's main city.

4. Results

■ **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 entry observed in the market. The table also reports 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. I compute the response to a 1% increase in each exogenous variable above its average across all locations in the dataset on a location-by-location basis. The reported marginal effects represent the average response in probabilities across markets and locations.¹⁶

Most of the parameter estimates are of the anticipated sign. They are precisely estimated due to the large within-market variation in the location characteristics. In both specifications, population has a large and positive effect on payoffs, but this effect decreases significantly with distance. A 1% increase in the location's population implies, for example, approximately a 3% increase in the likelihood of choosing this location. In contrast, a 1% increase in the most distant population in the market area increases the likelihood of choosing the location by only 1% to 2%. Business density has a negative effect on profits, while average per capita income has the expected positive effect, both in the location itself and 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 a higher effect on profits than per capita income levels in the most distant locations. The marginal effects of the latter on the likelihood of choosing a location are approximately .9%, about half of the marginal effect of average per capita income levels in the closer distance bands.

As expected, the presence of competitors has a negative effect on payoffs. This effect decreases significantly with distance, however. For example, an additional competitor within half a mile of a firm's location has a payoff effect that is approximately 70% stronger than that of an additional competitor within a half-mile to three miles, which in turn has a 52% to 66% stronger effect than a rival 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 rivals.

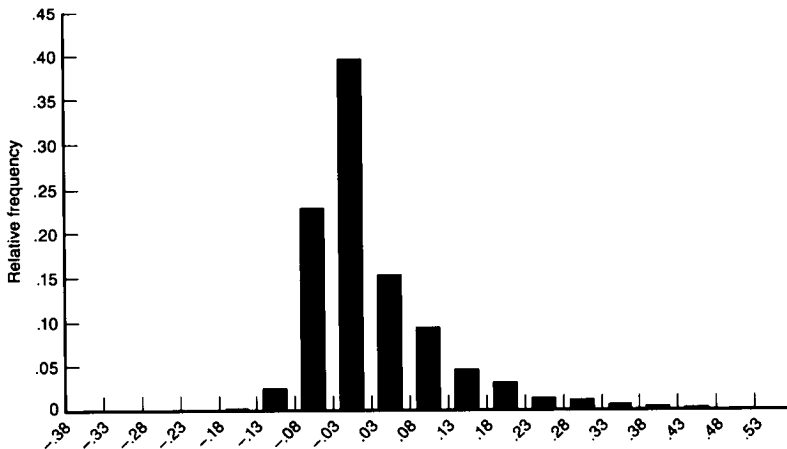
The impact of changing the size of the potential entrant pool is most pronounced in the estimate of the mean of the market-level effect, μ . If only a small fraction of the potential entrants enters the market, the market-level effect ξ adjusts relative to the outside option's mean zero profitability to reflect the revealed low attraction 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 case of the 50-firm potential entrant pool relative to the pool size set to twice the actual number of entrants. With 50 potential entrants, in most markets the fraction of firms that enters 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, there is a larger number of draws from the random payoff component ε . 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 affect the parameter estimates of the observed location-specific payoff determinants. The results show, however, that the estimates of these parameters are 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 third column of Table 4. The mode of the distribution is slightly below zero. The

¹⁶ Marginal effects cannot be computed for the endogenously determined expected number of competitors in the chosen location and in surrounding locations.

FIGURE 3
DISTRIBUTION OF PREDICTION ERRORS, LOCATION-CHOICE PROBABILITIES



distribution is skewed because 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, some prediction errors 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 .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 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.

□ **Illustration of results.** The estimated parameters of the location-choice model suggest that the competitive interaction between firms is strong at the level of the neighborhood. The lessening of competitive effects with distance indicates that firms exploit the geographic dispersion in demand to avoid competition with more distant rivals. Accordingly, we would expect more 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.

Physical location in clearly delimited markets permits investigation of the effect of individual features of the product-type space, such as 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 are not as easily observable. Other exercises of interest might include an investigation of the effect of regulation that

TABLE 4 Parameter Estimates, Entry and Location-Choice Model

Variable	Potential Entrant Pool			
	2 × Total Entrants		50 Firms	
	Coefficient (Standard Error)	Marginal Effect	Coefficient (Standard Error)	Marginal Effect
Population ₀ (000)	1.8191 (.1534)	.0333	2.1258 (.1764)	.0393
Population ₁ (000)	1.3109 (.1200)	.0236	1.7349 (.1498)	.0314
Population ₂ (000)	.6070 (.1192)	.0121	1.1348 (.1486)	.0227
Business Density	-.8077 (.1458)	-.0155	-.8889 (.1477)	-.0173
Average Per Capita Income ₀ (0000)	.9309 (.1136)	.0180	1.0380 (.1233)	.0204
Average Per Capita Income ₁ (0000)	1.0081 (.2081)	.0193	.9188 (.2043)	.0178
Average Per Capita Income ₂ (0000)	.4851 (.2512)	.0092	.4884 (.2601)	.0094
γ_0	-3.4520 (.3111)		-3.3853 (.3266)	
γ_1	-1.0103 (.0745)		-1.0087 (.0923)	
γ_2	-.3448 (.0738)		-.4870 (.0934)	
σ	3.5829 (.3110)		4.6760 (.4316)	
μ	-2.8764 (1.3425)		-7.0364 (1.5801)	

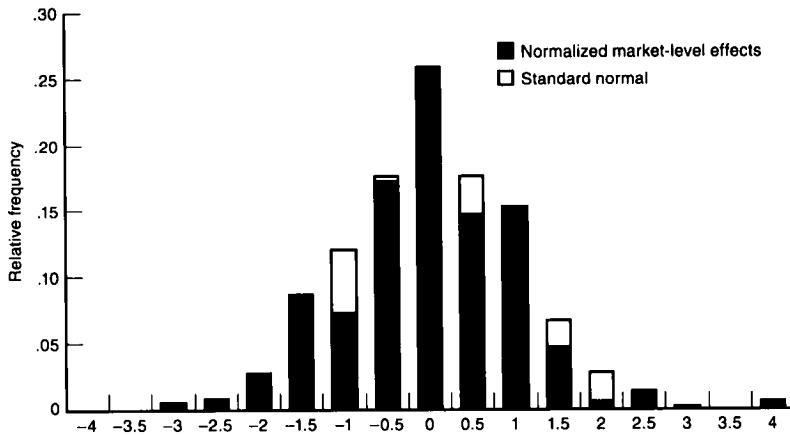
Note: Results based on 1999 demographic and firm data. Subscript 0 denotes the immediately adjacent locations to the chosen tract, within .5 miles in distance; subscript 1 denotes tracts at .5 to 3 miles in distance from the chosen tract; and subscript 2 denotes tracts at more than 3 miles distance from 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 are the estimates of the parameters of the distribution of ξ .

restricts the extent of product differentiation between firms. In the context of spatial differentiation, such regulation often takes the form of zoning ordinances. Similar examples from other contexts include licensing or minimum safety standards.

To isolate the effect 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, but its population increases as well. Comparing predicted entry patterns across the sample markets that vary in size 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. To separate the contribution of these factors, 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. To do so, I take one of the smallest sample cities, Jamestown, New York, with twelve Census tract locations, and arti-

FIGURE 4
DISTRIBUTION OF MARKET-LEVEL EFFECTS



cially increase its 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, only the population density in the twelve locations increases, leaving the area that the city occupies unchanged.

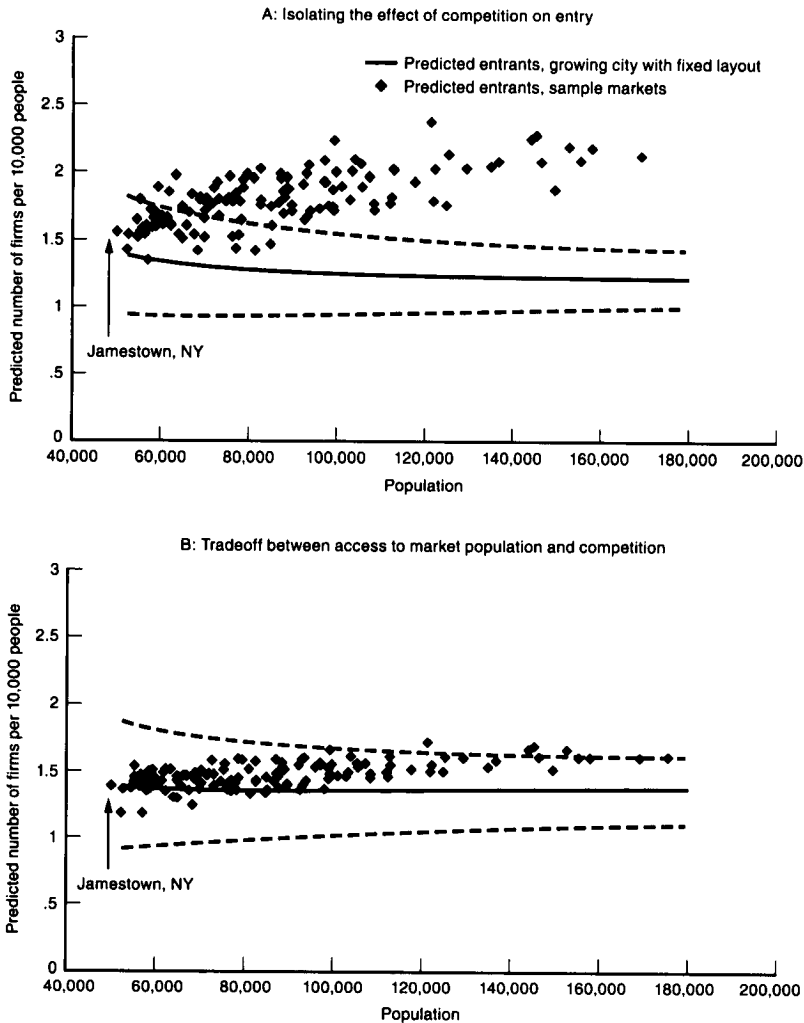
Predicted entry under this city expansion path is then contrasted with entry that would occur were the city to grow both in population and area. While it is difficult to simulate how Jamestown would expand if it grew, the cross-section of sample markets can be used as a proxy for this growth path. The sample markets are suitable for this purpose, since they span a range of market sizes from Jamestown at the lower end with a population of 52,583 to larger markets such as Fort Collins, Colorado, with a population of 178,070. Furthermore, the larger cities in the sample cover larger areas than the smaller cities and can thus represent how Jamestown might look were it to grow in population. 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 their values 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 comes from intensified competition. If the market area does not grow with population, firms cannot spread out in space any further, decreasing the incentive for additional entry. The second countervailing effect arises because 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. Increased access to nearby consumers thus increases the incentive for additional entry into the growing city with fixed spaces, relative to a city that grows in both population and space. The net effect of these two forces determines entry into a city with a fixed market area in comparison with 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 per 10,000 people into the actual sample markets. The solid line represents average predicted entry into the expanding Jamestown market, while the dotted lines denote the corresponding 95% confidence bands for entry.¹⁷ The model predicts an approximately linear relationship between market size

¹⁷ 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.

FIGURE 5
THE ROLE OF SPATIAL DISPERSION ON ENTRY



and equilibrium number of firms for the specific range of video store market structures covered by the sample. As a result, per capita entry is near constant or slightly declining.

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 distance band of .5 to 3 miles. 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 local 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 per 10,000 people from 1.4 to 2.1 stores. This increase in per capita entry relative to per capita entry in the case of a fixed area thus represents the contribution that the increased scope for spatial differentiation among firms makes to the number of firms that can profitably coexist in growing markets.

As a city grows in space, customers at one end of the city are less likely to frequent a store at the other end, so 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 local 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 in the immediate neighborhood increases payoffs, but this contribution falls short of the effect of increased competition on payoffs. As a result, predicted per capita entry into the sample markets on average exceeds predicted per capita entry into the growing Jamestown market with a fixed area. The competition effect thus dominates. On net, however, allowing the area of the city to increase with its population does not lead to 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 local.

5. Conclusion

■ In this article I present a framework for incorporating endogenous product-type choices into firms' entry decisions. I measure the subsequent impact on market structure for the video rental industry. The usefulness of the model lies in the fact that many markets have differentiated products. My results illustrate the incentives for spatial differentiation in the video rental industry, where location is a major source of product differentiation. Firms' abilities to capture local market power increase significantly with the size of the market space. As market area grows, however, population spreads out as well, limiting the benefits of spatial differentiation. In this particular application, 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 increases in market power offset the associated decrease in local demand only slightly.

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

I address the modelling difficulties that arise due to the complexity of multiple agents' product-differentiation decisions by using an imperfect-information framework in a static profit-maximizing context. As in equilibrium models of social interactions,¹⁸ firms endogenously sort into different locations depending on their expectation of the intensity of competition. The incomplete-information framework differs from its complete-information counterpart, including easier derivation of 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 *ex ante* assessment of the distribution of rival locations may not be the best response to its competitors' actual choice realizations. Allowing *ex post* regret corresponds better to real-world environments and decision making by firms; for example, when entry costs are sunk and rivals' profits are difficult to observe.

¹⁸ 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 (2005) and Timmins (2006).

In my model, market structure is the equilibrium outcome of firms' simultaneous location choices based on single-period payoff comparisons. The static model is assumed to approximate the repeated firm interaction that characterizes the evolution of an industry; in part because dynamic models of industry evolution such as Ericson and Pakes (1995) are computationally complex and difficult to estimate for markets with many product types and many firms. Recent work on empirically implementing dynamic games may shed light on the validity of this assumption.

Appendix

■ The location-choice model results in a set of equilibrium location conjectures defined by equation (7), 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 (7) 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,

$$\Psi(\mathbf{p}, \mathbf{X}, \mathcal{E}) = \mathbf{p} - F(\mathbf{p}, \mathbf{X}, \mathcal{E}) = 0, \tag{A1}$$

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

1. $\frac{\partial \Psi_\ell}{\partial p_\ell} > 0$
2. $\left| \frac{\partial \Psi_\ell}{\partial p_\ell} \right| \geq \sum_{k \neq \ell} \left| \frac{\partial \Psi_\ell}{\partial p_k} \right|$.

Consider first the example of a 2×2 city that allows for spatially differentiated competition by letting the effect of competitors in a given location differ from that of competitors in the remaining three locations. Expected profits for this example are thus given by

$$\mathbb{E}[\Pi_{f_\ell}] = \xi + \mathbf{X}_\ell \beta + \gamma_0(1 + (\mathcal{E} - 1)p_\ell) + \gamma_1(\mathcal{E} - 1) \sum_{k \neq \ell} p_k + \varepsilon_{f_\ell} \tag{A2}$$

for $\ell, k = 1, \dots, 4$. Normalizing $p_4 = 1 - (p_1 + p_2 + p_3)$, the matrix of partial derivatives for this specific example contains the elements

$$\begin{aligned} \frac{\partial \Psi_\ell}{\partial p_\ell} &= 1 - (\mathcal{E} - 1)p_\ell(\gamma_0 - \gamma_1)(1 - p_\ell + p_4) \\ \frac{\partial \Psi_\ell}{\partial p_k} &= -(\mathcal{E} - 1)p_\ell(\gamma_0 - \gamma_1)(-p_k + p_4) \quad k \neq \ell, \ell, k = 1, 2, 3. \end{aligned} \tag{A3}$$

Assuming without loss of generality that $p_4 = \min(\mathbf{p})$, 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 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 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. Nonuniqueness arises in these scenarios in particular if there are only little or no differences in locations' demographic makeup to induce additional variation in payoffs across locations.

The profit function used in estimation has three distance bands 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

$$\begin{aligned} \frac{\partial \Psi_\ell}{\partial p_\ell} &= 1 - (\mathcal{E} - 1)p_\ell \\ &\times \left[(\mathbb{I}_{\ell\mathcal{L}}^1(\gamma_0 - \gamma_1) + \mathbb{I}_{\ell\mathcal{L}}^2(\gamma_0 - \gamma_2))(1 - p_\ell + p_{\mathcal{L}}) + (\gamma_1 - \gamma_2) \left(\sum_{k \neq \ell} (\mathbb{I}_{k\ell}^2(1 - \mathbb{I}_{k\mathcal{L}}^2) - \mathbb{I}_{k\ell}^1(1 - \mathbb{I}_{k\mathcal{L}}^1))p_k \right) \right] \end{aligned} \tag{A4}$$

$$\frac{\partial \Psi_\ell}{\partial p_k} = -(\mathcal{E} - 1)p_\ell \times \left[(\mathbb{I}_{k\mathcal{L}}^1(\gamma_0 - \gamma_1) + \mathbb{I}_{k\mathcal{L}}^2(\gamma_0 - \gamma_2))(-p_k + p_\mathcal{L}) + (\gamma_1 - \gamma_2) \left(\mathbb{I}_{k\ell}^1 + \sum_{j \neq k} (\mathbb{I}_{jk}^2(1 - \mathbb{I}_{j\mathcal{L}}^2) - \mathbb{I}_{jk}^1(1 - \mathbb{I}_{j\mathcal{L}}^1))p_j \right) \right] \quad (\text{A5})$$

$$k \neq \ell, \ell, k = 1, \dots, \mathcal{L} - 1,$$

where $p_\mathcal{L}$ has been normalized to $1 - \sum_{\ell \neq \mathcal{L}} p_\ell$ and, as before, $\mathbb{I}_{k\ell}^b = 1$ if $\mathcal{D}_b \leq d_{k\ell} < \mathcal{D}_{b+1}$ 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 \mathbf{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 dataset of markets that differ in the amount of variation in demographic attributes.¹⁹ 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 (7). 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.

Expressions (A4) and (A5) 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. *Video Store Magazine*. Santa Ana, Calif., various issues.
- AGUIRREGABIRIA, V. AND MIRA, P. "Sequential Estimation of Dynamic Discrete Games." Mimeo, Boston University, 2006.
- AHN, H. AND MANSKI, C. "Distribution Theory for the Analysis of Binary Choice Under Uncertainty with Nonparametric Estimation of Expectations." *Journal of Econometrics*, Vol. 56 (1993), pp. 291–321.
- ANDREWS, D., BERRY, S., AND JIA, P. "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Locations." Mimeo, Yale University, 2004.
- AUGEREAU, A., GREENSTEIN, S., AND RYSMAN, M. "Coordination Versus Differentiation in the Adoption of High-Speed Modems." *RAND Journal of Economics*, forthcoming.
- BAYER, P., McMILLAN, R., AND RUEBEN, K. "An Equilibrium Model of Sorting in an Urban Housing Market." Mimeo, Yale University, 2005.
- BERRY, S. "Estimation of a Model of Entry in the Airline Industry." *Econometrica*, Vol. 60 (1992), pp. 889–917.
- . "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics*, Vol. 25 (1994), pp. 242–262.
- AND WALDFOGEL, J. "Free Entry and Social Inefficiency in Radio Broadcasting." *RAND Journal of Economics*, Vol. 30 (1999), pp. 397–420.
- , LEVINSOHN, J., AND PAKES, A. "Automobile Prices in Market Equilibrium." *Econometrica*, Vol. 63 (1995), pp. 841–890.
- BRESNAHAN, T.F. AND REISS, P. "Entry in Monopoly Markets." *Review of Economic Studies*, Vol. 57 (1990), pp. 531–553.

¹⁹ 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.

- AND ———. "Entry and Competition in Concentrated Markets." *Journal of Political Economy*, Vol. 99 (1991), pp. 977–1009.
- BROCK, W. AND DURLAUF, S. "Interactions-Based Models." In J. Heckman and E. Leamer, eds., *Handbook of Econometrics*. Amsterdam: North-Holland, 2001.
- CILIBERTO, F. AND TAMER, E. "Market Structure and Multiple Equilibria in Airline Markets." Mimeo, North Carolina State University, 2006.
- COTTERILL, R.W. AND HALLER, L.E. "Barrier and Queue Effects: A Study of Leading U.S. Supermarket Chain Entry Patterns." *Journal of Industrial Economics*, Vol. 40 (1992), pp. 427–440.
- DAVIS, P. "Spatial Competition in Retail Markets: Movie Theaters." *RAND Journal of Economics*, forthcoming.
- EINAV, L. "Not All Rivals Look Alike: Estimating an Equilibrium Model of the Release Date Timing Game." Mimeo, Stanford University, 2003.
- ELLRICKSON, P. "Supermarkets as a Natural Oligopoly." Mimeo, Duke University, 2004.
- ERICSON, R. AND PAKES, A. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, Vol. 62 (1995), pp. 53–82.
- HASTING'S BOOK, MUSIC AND VIDEO, INC. *1998 Annual Report*. Amarillo, Tex., 1998.
- HOTELLING, H. "Stability in Competition." *Economic Journal*, Vol. 39 (1929), pp. 41–57.
- LANCASTER, K. "A New Approach to Consumer Theory." *Journal of Political Economy*, Vol. 74 (1966), pp. 132–157.
- . *Variety, Equity and Efficiency: Product Variety in an Industrial Society*. New York: Columbia University Press, 1979.
- MAZZEO, M. "Product Choice and Oligopoly Market Structure." *RAND Journal of Economics*, Vol. 33 (2002), pp. 221–242.
- MCKELVEY, R.D. AND PALFREY, T.R. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior*, Vol. 10 (1995), pp. 6–38.
- PAKES, A., OSTROVSKY, M., AND BERRY, S. "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)." Mimeo, Harvard University, 2005.
- PESENDORFER, M. AND SCHMIDT-DEGLER, P. "Identification and Estimation of Dynamic Games." NBER Working Paper no. 9726, 2003.
- REISS, P. "Empirical Models of Discrete Strategic Choices." *American Economic Review*, Vol. 86 (1996), pp. 421–426.
- RUST, J. "Estimation of Dynamic Structural Models." In C. Sims., ed., *Advances in Econometrics: Sixth World Congress*. New York: Cambridge University Press, 1994.
- SEIM, K. "Spatial Differentiation and Market Structure: The Video Retail Industry." Ph.D. dissertation, Department of Economics, Yale University, 2001.
- SINNOTT, R.W. "Virtues of the Haversine." *Sky and Telescope*, Vol. 68 (1984), p. 159.
- STAVINS, J. "Model Entry and Exit in a Differentiated-Product Industry: The Personal Computer Market." *Review of Economics and Statistics*, Vol. 77 (1995), pp. 571–584.
- SUTTON, J. *Sunk Costs and Market Structure*. Cambridge, Mass.: MIT Press, 1991.
- . *Technology and Market Structure*. Cambridge, Mass.: MIT Press, 1998.
- TAMER, E. "Incomplete Simultaneous Discrete Response Models with Multiple Equilibria." *Review of Economic Studies*, Vol. 70 (2003), pp. 147–165.
- THOMADSEN, R. "The Effects of Ownership Structure on Prices in Geographically Differentiated Industries." *RAND Journal of Economics*, Vol. 36 (2005), pp. 908–929.
- TIMMINS, C. "Estimating Spatial Differences in the Brazilian Cost-of-Living with Household Location Choices." *Journal of Development Economics*, Vol. 80 (2006), pp. 59–83.
- TOIVANEN, O. AND WATERSON, M. "Empirical Research on Discrete Choice Game Theory Models of Entry: An Illustration." *European Economic Review*, Vol. 44 (2000), pp. 985–992.
- AND ———. "Market Structure and Entry: Where's the Beef?" *RAND Journal of Economics*, Vol. 36 (2005), pp. 680–699.
- VIDEO SOFTWARE DEALERS' ASSOCIATION. *An Annual Report on the Home Video Market 1998*. Encino, Calif., 1998.
- WATSON, R. "Product Variety and Competition in the Retail Market for Eyeglasses." Mimeo, University of Texas at Austin, 2005.