Quant Mark Econ (2013) 11:83–116
DOI 10.1007/s11129-012-9131-x

Does reducing spatial differentiation increase product differentiation? Effects of zoning on retail entry and format variety

Sumon Datta · K. Sudhir

Received: 8 December 2010 / Accepted: 20 September 2012 / Published online: 13 October 2012

Abstract This paper investigates the impact of spatial zoning restrictions on retail market outcomes. We estimate a structural model of entry, location and format choice across a large number of markets in the presence of zoning restrictions. The paper contributes to the literature in three ways: First, the paper demonstrates that the omission of zoning restrictions in the extant literature on entry and location choice leads to biased estimates of the factors affecting market potential and competitive intensity. Second, the cross-market variations in zoning regulations helps us test and provide evidence for the theory that constraints on spatial differentiation will lead to greater product differentiation. Finally, we provide qualitative insight on how zoning impacts retail entry and format variety; in particular we evaluate the impact of prototypical zoning arrangements such as "centralized," "neighborhood," and "outskirt" zoning on entry and format variety.

Keywords Product variety · Zoning · Entry · Location choice · Retail competition · Discrete games · Multiple equilibria · Structural modeling

JEL Classification C5 · C70 · L13 · L81 · M31

> "*With a city entirely zoned they [realtors] could assure purchasers of residential property that their neighborhoods would never be encroached upon by business, while on the other hand, zoning would give business property a touch*

The paper is based on an essay from the first author's dissertation at the Yale School of Management. We thank Ahmed Khwaja, Jiwoong Shin and participants at the 2011 SICS conference, 2011 Marketing in Israel Conference, 2011 Marketing Science Conference, seminars in the Marketing departments at University of Connecticut, Fudan, Texas A&M, and Ph.D. student workshops in the Marketing departments at Yale School of Management, Purdue University and University of Michigan for their comments.

S. Datta (✉)
Krannert School of Management, Purdue University, 403 W. State Street, West Lafayette, IN 47907, USA
e-mail: sdatta@purdue.edu

K. Sudhir
Yale School of Management, 135 Prospect St, PO Box 208200, New Haven, CT 06520, USA
e-mail: k.sudhir@yale.edu

of monopoly value. Accordingly, the signs went up on vacant lots: "Zoned for business," or "Zoned for apartments" with the definite implication that such action on the part of the public authorities had resulted in giving the property a higher and more assured value than it would otherwise have." Munro (1931, p. 203)

1 Introduction

Zoning is a device for land use planning used by local governments, and refers to the practice of designating permitted uses of land based on mapped zones within their jurisdictions which separate one set of land uses from another. Broadly, land is zoned for residential, industrial, commercial, agricultural, forests, open spaces and recreational purposes among others. Typically there are also finer regulations on the types of residences, industries and commercial ventures that are allowed on particular parcels of land. The practice of zoning in modern days began in the 1860's in Germany (Ladd 1990), and was widely embraced in the United States during the 1920's, with New York City passing its first zoning ordinance in 1916 (Fischel 2004). While there has been much research, discussions and debate in the urban planning literature about the motivations and effects of residential zoning (see Chung (1994) and Pogodzinski (1991) for reviews), there has been limited research on how zoning impacts market outcomes. For example, how do changes in zoning impact retailer entry and choice of formats due to their impact on retail competition and profits in equilibrium?

The paper estimates a static, structural simultaneous move game of endogenous entry, location and format choice across a large number of markets within the United States, taking into account the various local zoning restrictions on commercial entry. We use our analysis to help answer substantive, econometric and theoretical questions of interest related to spatial zoning. First, the paper introduces the zoning dimension to the by-now extensive literature on entry (e.g., Bresnahan and Reiss 1991; Mazzeo 2002; Vitorino 2010; Zhu et al. 2009; Ciliberto and Tamer 2009), and location choice (e.g., Seim 2006; Orhun 2012; Watson 2009; Zhu and Singh 2009). Having access to zoning data helps us answer an econometric question with important substantive implications. Specifically, does the omission of zoning restrictions in the extant literature on entry and location choice bias estimates of factors affecting market potential and competitive intensity and if so, by how much? Zoning is a constraint with a negative impact on that location to generate profits. When zoning is omitted, this negative unobservable is absorbed in the error term.[1] To the extent that zoning is correlated with market characteristics such as population and income that impact retailer profits, this omission will lead to biases in the estimated coefficients for these variables. For example, if locations with higher incomes systematically have tougher zoning restrictions that restrict entry, incomes and the unobservables will be negatively correlated, and the income coefficient will be biased towards zero (i.e., the effect of income on profits will be underestimated, because profitable high income locations will appear

[1] For example, Orhun (2012) considers the importance of allowing for location specific unobservables to potentially control for omitted location characteristics such as zoning, public transportation and major road intersections.

unprofitable due to lack of entry due to zoning restrictions in those locations). The direction of the bias for competition effects is harder to predict ex ante because the effect of zoning on competitor entry depends on the correlation between zoning restrictions and market characteristics like income and population.[2] The bias in estimates is not merely an econometric issue; given the difficulties associated with assembling zoning data, understanding the magnitude of the bias due to the omission of zoning restrictions can be valuable in guiding whether policy makers, firms and researchers need to invest in collecting zoning data in making decisions or recommendations.

Second, the paper contributes to the theoretical literature on the link between spatial and product differentiation. Theory predicts that constraints on spatial differentiation will lead to greater product differentiation; for example, in the context of the Internet, Kuksov (2004) conjectured that firms will respond to the inability to spatially differentiate on the Internet with greater product differentiation. Bar-Isaac et al. (2009) elaborate on this argument of endogenous differentiation to explain the long-tail effect of greater product variety in Internet retail environments. However the theory has not faced empirical scrutiny. As zoning restrictions affect the ability of retailers to spatially differentiate, we exploit cross-market variations in spatial zoning restrictions for retailers to empirically test the validity of the theoretical prediction. Specifically, we test whether tighter zoning restrictions lead to greater format variety. In the context of retailing, if zoning prevents spatial differentiation, retailers will differentiate more on retail formats. For example, in food retailing, retailers can offer many formats: supermarkets, supercenters, convenience stores, mass-merchandising etc. Hence, if theory were true, tighter zoning will increase format variety without necessarily reducing the number of stores.

Finally, it provides substantive insight for citizens and local regulators who decide on zoning regulations, and firms who have to assess the equilibrium impact of zoning regulations. We use the estimates of the structural model to perform counterfactual simulations on how zoning regulations impacts entry, location and format choice. Specifically, we assess how certain "prototype" zoning approaches such as "centralized," "neighborhood" or "outskirt" zoning affect retail entry and retail format mix.

The paper leverages on a method introduced in Datta and Sudhir (2012) to obtain zoning data from a publicly available digital dataset called National Land Cover Dataset (NLCD). In spite of the importance of zoning in retail entry and location decisions, extant research has ignored the issue primarily because of the lack of easily available zoning data across a large number of markets. We use NLCD in conjunction with Geographic Information System tools such as *ArcGIS* and *Google Earth* to recover zoning data in any number of markets across the entire U.S.

We use maximum likelihood estimation for estimation of the static discrete game.[3] Some well-known methodological challenges in estimating discrete games include the possibility of multiple equilibria in the model, multiple equilibria in the data, and slow convergence or potential non-convergence of the MLE estimation algorithm. We use recent innovations in the literature to address these issues. We use the nested pseudo likelihood (NPL) approach to address the equilibrium selection challenge in

[2] See Section 4.1 for a more elaborate discussion of bias in competition effects.

[3] Alternatives to likelihood based approaches include method of moments (Thomadsen 2005; Draganska et al. 2009), minimum distance or asymptotic least square estimators (Pakes et al. 2007; Bajari et al. 2007; Pesendorfer and Schmidt-Dengler 2008) and maximum score estimators (Fox and Bajari 2010; Fox 2007; Ellickson et al. 2010).

the face of multiple equilibria in the model by selecting the equilibrium most consistent with the conditional choice probabilities (CCP) in the data. To address the challenge of equilibrium selection in the presence of multiple equilibria in the data, we combine a "parallel NPL" procedure, which intuitively involves starting from different starting values of CCP with a genetic algorithm approach which ensures we search over a large space of equilibria as suggested by Aguirregabiria and Mira (2005). Finally, to speed convergence, we use a transformed contraction mapping suggested by Kasahara and Shimotsu (2008).

Besides the literature on entry and location games, there is a nascent and contemporaneous empirical literature that has begun to explore the relationships between market structure and zoning or land use regulations. Suzuki (2010) and Nishida (2010) study the endogenous market entry decisions of firms, controlling for market level land use regulations that involve additional investment or time on behalf of a firm in order to obtain the permission to enter a market.[4] Not surprisingly, they find that such land use regulations can be anti-competitive by acting as a barrier for entry through higher entry costs. Ridley et al. (2010) use zoning data for 15 municipalities in the Minneapolis-St. Paul area to study how zoning impacts the number of rivals, prices of the central retailer and the average distance from rivals. Through reduced form regressions, they show that in the geographical area around a central retailer, the fraction of area that is zoned for commercial use has a positive correlation with the number of rivals, the prices of the central retailer, and its average distance from rivals. In sum their results support the hypothesis that tighter commercial zoning leads to fewer rivals, but still overall a firm faces more price competition because of zoning-forced spatial proximity. None of these papers tackle the issue of format variety, which can moderate the effects on entry and competition.

Our key findings are as follows: First, we find significant biases in the estimates for factors affecting market potential and competitive intensity when zoning is not accounted for. Therefore future empirical work on entry and location choice needs to incorporate zoning restrictions in their analysis. Second, zoning restrictions do reduce entry, but over small ranges of restrictions, firms respond by increasing format variety without reducing entry. This suggests that if one does not take into account format responses, one might see weak linkages between zoning restrictions and entry and potentially conclude that zoning has limited impact on retail entry decisions. Substantively, we find that different prototypical arrangements like centralized, neighborhood and outskirt zoning can lead to different retail structures in terms of both the number and type of stores in a market. Outskirt zoning leads to more homogeneous formats, while centralized zoning leads to more format variety. Finally, we demonstrate empirical evidence to the theoretical conjectures that firms indeed respond to tightened spatial differentiation (or inability to spatially differentiate) through greater product (in our case format) differentiation.

The rest of this paper is organized as follows: Section 2 describes the model and estimation strategy. Section 3 describes the entry and location choice data and our

[4] Suzuki (2010) uses seven indices to measure the stringency of land use (zoning) regulations in 60 Texas counties to study the entry decisions of mid-scale chain hotels in those counties. Nishida (2010) studies the entry decisions of convenience store chains in Okinawa, Japan. In his application, an entire market is counted as a zoned market if retailers are required to obtain development permission from the government in order to enter the market.

approach to obtaining spatial zoning data. Section 4 discusses the potential sources of estimation bias followed by a discussion of the estimates of the model and Section 5 presents the results of counterfactual simulations. Section 6 concludes with a brief summary of the findings and the limitations of this research.

2 Model and estimation strategy

2.1 Model of strategic entry, location and store format choice

We model the entry, location and store format choice as a two-stage game in which the firms first make entry decisions and then the location and format choices. More precisely, in the first stage, each firm, i decides whether or not to enter a market m ($m=1, 2,\ldots, M$); subsequently, in the second stage the entering firms simultaneously choose their respective store type or format, f ($f=1,2,\ldots, F$), and store location within the market.

For the purposes of illustration, imagine a square city with a grid of L^m discrete blocks or 'locations' (Fig. 1(a)). Firm i's payoff at each location, l ($l=1, 2,\ldots, L^m$), is modeled as a function of the endogenous choice of format type, f, the market characteristics at the location, x_l, the actions (entry, location and format choices) of all firms, $a=(a_i, a_{-i})$, and an idiosyncratic profit shock, ε_{ifl}, which is the firm's private information and is known to rivals (and the researcher) only in distribution:

$$\pi^m_{ifl}(a_i) = \Pi^m_f(x_l, a) + \varepsilon_{ifl} \tag{1}$$

In this incomplete information setup, a firm cannot exactly predict rivals' actions but it has rational beliefs about their strategies. For example, suppose firms are homogeneous, then each firm will make its decision based on its belief about the number of firms that would enter the market, N^m, and its belief that an entering rival will choose a particular location as represented by a vector of conditional location choice probabilities (CCP), $\overline{P}^m \left(\overline{P}^m = \{p_1, p_2, \ldots, p_{L^m}\}\right)$. For instance, the firm may

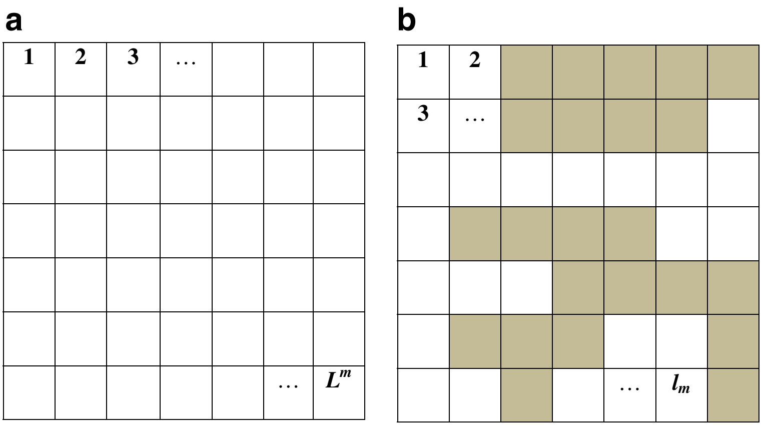


Fig. 1 **a** An illustrative square market with the geographical space discretized into square blocks or '*locations*'. **b** Due to zoning regulations, firms can only choose among '*potential retail location*' (*Area in white*)

have a belief that a rival, conditional on entry, will choose location 'j' with probability p_j. Hence, for homogeneous firms the expected profit at location l can be written as (after dropping subscript 'f' for format):

$$E\left[\pi_{il}^{m}(a_i)\right] = \Pi^{m}\left(x_l, \left(N^{m}, \overline{P}^{m}\right)\right) + \varepsilon_{il} \tag{2}$$

In extant models, firms are allowed to consider all L^m locations in the market so that each location has some positive probability of being chosen by a firm. However, since firms are not allowed to set up stores in residential locations, we use our zoning data to exclude such locations and concentrate only on a subset of *potential retail locations*, $l=\{1, 2, \ldots, l_m\}$ (Fig. 1(b)). Hence, retailers' ability to differentiate spatially, and the level of consumers' search cost, is driven by the l_m zoned locations.

Like Seim (2006), we divide the area around a location into concentric circles or distance bands. All consumers on a distance band b ($b=1, 2, \ldots, B$) around location l are assumed to have the same effect on the firm's profit. Also, all rivals of a particular format type that are on distance band b are assumed to have the same competitive effect on the firm. Hence, Eq. (1) can be expanded as follows:

$$\pi_{ifl}^{m} = \sum_{b=1}^{B} \alpha_{fb} \cdot x_{lb} + \sum_{b=1}^{B} \beta_{f-fb} \cdot N_{tlb}^{m} + \sum_{b=1}^{B} \sum_{f' \neq f} \beta_{f'-fb} \cdot N_{f'lb}^{m} + \xi^{m} + \varepsilon_{ifl} \tag{3}$$

where, x_{lb} is a vector of location characteristics like population and per capita income in distance band b around location l. The impact of these location characteristics (α_{fb}) on profits is allowed to be format-specific as denoted by the subscript 'f'. The second term on the right hand side of Eq. (3) is the *intra-format* competition effect where N_{flb}^{m} is the number of rivals in distance band b that have the same format, f, as the focal firm, i, and β_{f-fb} is the competitive effect of one such rival. Next, is the *inter-format* competition effect where $N_{f'lb}^{m}$ is the number of rivals with format, f' that is different from the focal firm ($f' \neq f$), and $\beta_{f'-fb}$ is the competitive effect of one such f'-format rival. We expect β_{f-fb} and $\beta_{f'-fb}$ to fall in magnitude with increasing b, reflecting lower competition between rivals at greater distances, reflecting the benefit of spatial differentiation for firm profit. We also expect $\beta_{f-fb} > \beta_{f'-fb}$ to reflect that intra-format competition will be lower than inter-format competition, reflecting the benefit of format differentiation on firm profit. Finally, ξ^m captures the unobserved attractiveness of the market that cannot be explained by the observable market characteristics. It would include market characteristics that are unobserved by the researcher but that are common knowledge for firms when they make their decision.[5]

Firms have rational expectations about rivals' strategies in equilibrium so that firm i expects a particular number of f-format (f'-format) rivals in distance band b,

[5] We assume that there are no location-specific profit unobservables that are common knowledge for firms at the time of entry but unobservable to the researcher. As we do not have data on major road intersections, and high rents or tax rates (that are independent of the observed variables such as population), this could lead to potentially biased estimates. Orhun (2012) considers the importance of allowing for such location specific unobservables.

$E\left[N_{f'lb}^m\right]\left(E\left[N_{flb}^m\right]\right)$. Hence, the firm can form expectations about its profit at each location as follows:

$$E\left[\pi_{ifl}^m\right] = \sum_{b=1}^{B} \alpha_{fb} \cdot x_{lb}^m + \sum_{b=1}^{B} \beta_{f-fb} \cdot E\left[N_{flb}^m\right] + \sum_{b=1}^{B} \sum_{f' \neq f} \beta_{f'-fb} \cdot E\left[N_{f'lb}^m\right] + \xi^m + \varepsilon_{ifl} \quad (4.1)$$

where,

$$E[N_{flb}^m] = (N^m - 1) \cdot \sum_{j \in \ell_{lb}} p_{fj} \quad (4.2)$$

$$E[N_{f'lb}^m] = (N^m - 1) \cdot \sum_{j \in \ell_{lb}} p_{f'j} \quad (4.3)$$

Here, ℓ_{lb} is the set of locations in distance band b around location l. p_{fj} ($p_{f'j}$) is a conditional choice probability for store format f (f') and location j firm. It represents the focal firm's belief that a rival firm will open an f-format (f'-format) store in location j when a total of N^m firms enter the market. Hence, corresponding to f-format (f'-format) firms we will have a vector of l_m conditional location choice probabilities, $\overline{P}_f^m = \{p_{f1}, p_{f2}, \ldots, p_{fl_m}\}$ $\left(\overline{P}_{f'}^m = \{p_{f'1}, p_{f'2}, \ldots, p_{f'l_m}\}\right)$. Hence, for each market, we essentially have a matrix of l_m X F conditional format and location choice probabilities (conditional on N^m firms entering the market), $\overline{P}^m = \left[\overline{P}_1^m, \overline{P}_2^m, \ldots \overline{P}_F^m\right]$.

Now, analogous to Eq. (2), we can rewrite Eq. (4.1) in terms of the total number of entrants, N^m, a matrix of firm's beliefs about rivals' conditional location choice probabilities, $\overline{P}^m$, and a set of model parameters, θ:

$$E\left[\pi_{ifl}^m\right] = \widehat{\pi}_{ifl}\left(x_{lb}^m, N^m, \overline{P}^m; \theta\right) + \xi^m + \varepsilon_{ifl} \quad (5)$$

Note that ξ^m is common for all store formats as well as across all locations within a market, and therefore does not influence the location choice after firm i has decided to enter the market. Thus, if we assume that the private information shock, ε_{ifl}, has a i.i.d. type 1 extreme value distribution, then the conditional choice probability (conditional on entry) that an entering firm chooses to open a f-format store in location l is given by the logit form:

$$\psi_{fl}\left(x_{lb}^m, N^m, \overline{P}^m; \theta\right) = \frac{\exp\left[\widehat{\pi}_{ifl}\left(x_{lb}^m, N^m, \overline{P}^m; \theta\right)\right]}{\sum_{\varphi=1}^{F} \sum_{j=1}^{l_m} \exp\left[\widehat{\pi}_{i\varphi j}\left(x_{lb}^m, N^m, \overline{P}^m; \theta\right)\right]} \quad (6)$$

Next, we normalize the profit from not entering a market to zero so that the entry probability for a firm is given by the nested logit form:

$$p(Entry) = \frac{\exp(\xi^m) * \sum_{\varphi=1}^{F} \sum_{j=1}^{l_m} \exp\left[\widehat{\pi}_{i\varphi j}\left(x_{lb}^m, N^m, \overline{P}^m; \theta\right)\right]}{1 + \exp(\xi^m) * \sum_{\varphi=1}^{F} \sum_{j=1}^{l_m} \exp\left[\widehat{\pi}_{i\varphi j}\left(x_{lb}^m, N^m, \overline{P}^m; \theta\right)\right]} \quad (7)$$

Hence, if there are, say, R potential retail entrants then the expected total number of entrants in market m is given by:

$$N^m = R^* p(Entry) \tag{8}$$

Similar to Seim (2006, p. 625–626), we assume that the expected number of entrants in a market is exactly equal to the number of entrants observed in the data. That is, the market-level unobserved profit shock, ξ^m, is the lowest value for which the expected number of entrants predicted by the model, coincides with observed numbers in each market. By exogenously fixing R, and by observing the actual number of entrants, N^m, the unobserved market attractiveness parameter, ξ^m, can be therefore estimated using Eqs. (7) and (8):

$$\xi^m | N^m = \ln(N^m) - \ln(R - N^m) - \ln\left(\sum_{\varphi=1}^{F}\sum_{j=1}^{l_m} \exp\left[\widehat{\pi}_{i\varphi j}\left(x_{lb}^m, N^m, \overline{P}^m; \theta\right)\right]\right) \tag{9}$$

We assume that ξ^m is i.i.d. across markets, and follows a normal distribution, $N(\mu,\sigma^2)$. Thus the probability that a total of N^m firms enter the market is given by the p.d.f. of this normal distribution at the value obtained in Eq. (9). Note that the value of ξ^m adjusts to the size of R in relation to the outside option of no entry. Hence, although the size of R is not observed by the researcher, varying the size will have only a miniscule effect on our inferences about firms' strategies (See discussion in Seim (2006)).

Now, we can construct the likelihood with the constraint that each firm's beliefs about rivals' strategies must match those rivals' equilibrium strategies:

$$\begin{aligned} L\left(\overline{P}^m, \xi^m, \Theta\right) &= \prod_{m=1}^{M}\left[\underbrace{\left\{\prod_{f=1}^{F}\prod_{l=1}^{l_m}\left(\psi_{fl}\left(x_{lb}^m, N^m, \overline{P}^m; \theta\right)\right)^{I(fl)}\right\}}_{\textit{Format and Location Choice}} \underbrace{*\phi\left(\xi^m; \mu, \sigma^2\right)}_{\textit{Entry Choice}}\right] \\ s.t. \qquad p_{fl} &= \psi_{fl}(x_{lb}^m, N^m, \overline{P}^m; \theta), \qquad \forall l, \forall f, \forall m \end{aligned} \tag{10}$$

where, Θ is the set of all model parameters ($\Theta = \{\theta, \mu, \sigma^2\}$), and $I(fl)$ is an indicator that equals one if location l is chosen by a f-format firm, and is zero otherwise.

The constraint in Eq. (10) is a system of equations that defines firms' conditional location choice probabilities as the fixed point of a continuous mapping between firms' strategies and their rivals' strategies. As the conditional location choice probabilities in a market must add up to 1, by Brouwer's fixed point theorem, this system of equations has at least one solution or fixed point, $\overline{P}^* | \Theta$, for any value of Θ.[6]

[6] Like the rest of the entry and location literature, we assume household locations are exogenous to firm choices in this paper. To the extent, we analyze established markets in which populations and zoning restrictions are stable and there have been ample opportunities for firms to enter and exit and are in equilibrium before the period of analysis, our assumption might be reasonable. However, causality could be reversed in newly developing markets, where consumer growth could lead to new firm/format entry, which could in turn lead to certain types of consumer growth. We abstract away from these issues in our analysis. If new neighborhoods significantly dominate the data, the effect of market variables such as population on profits will be overestimated. We thank a reviewer for drawing our attention to this issue.

2.2 Estimation strategy

2.2.1 Simplifying restrictions

In the general model specification above, the number of model parameters increases exponentially with the number of format types (F) due to the inter-format and intra-format competition effects. The number of distance bands (B) around each location further explodes the number of parameters. Specifically, in our empirical application, we have six format types (F=6), and we consider five 1-mile width distance bands around each location (B=5). Therefore, the number of competition parameters alone is 180 (F^2*B=6*6*5). Given that we only have 100 sample markets (M) from which to estimate the model, we employ two restrictions to reduce the number of parameters to be estimated.

First, we assume that the competition effect between a pair of rival formats is symmetric. That is, for any distance band, b, and for two rivals with formats f and f', we assume $\beta_{f'-fb}=\beta_{f-f'b}$. Second, we assume a constant multiplier assumption on the effects across distance bands. Specifically, we assume that the impact of a factor on market potential at a particular distance band to be a constant multiplier of the impact of that factor in the first 0–1 mile distance band.[7] For instance, suppose the coefficients of population and per-capita income on store profit are denoted by $\alpha 1_{fb}$ and $\alpha 2_{fb}$, respectively; then the restriction implies:

$$\begin{array}{l} \alpha 1_{f2} = \alpha 1_{f1}\lambda_2;\ \alpha 1_{f3} = \alpha 1_{f1}\lambda_3\ ;\ \ldots;\alpha 1_{fB} = \alpha 1_{f1}\lambda_B \\ \alpha 2_{f2} = \alpha 2_{f1}\zeta_2;\ \alpha 2_{f3} = \alpha 2_{f1}\zeta_3\ ;\ \ldots;\alpha 2_{fB} = \alpha 2_{f1}\zeta_B \end{array} \tag{11.1}$$

Similarly for competition, we assume that:

$$\begin{array}{l} \beta_{f-f2} = \beta_{f-f1}\kappa_2;\beta_{f-f3} = \beta_{f-f1}\kappa_3;\ \ldots;\beta_{f-fB} = \beta_{f-f1}\kappa_B \\ \beta_{f'-f2} = \beta_{f'-f1}\kappa_2;\beta_{f'-f3} = \beta_{f'-f1}\kappa_3;\ \ldots;\beta_{f'-fB} = \beta_{f'-f1}\kappa_B \\ (\#\ \textit{competition related parameters} = (F^*(F+1)/2) + (B-1)) \end{array} \tag{11.2}$$

2.2.2 Multiple equilibria

The observed actions of firms in different markets in the data could potentially correspond to different equilibria. That is, there could be *multiple equilibria in the data*. Without any additional information about which equilibrium is favored for each market in the data, identification is obtained by assuming that every observation in the data comes from the same equilibrium. This assumption of *single equilibrium in the data* is common practice in the empirical literature on games of oligopoly competition, and it implicitly establishes the *equilibrium selection mechanism*. Under this condition, estimation involves finding the equilibrium solution, $(\overline{P}^*_{MLE}, \Theta^*_{MLE})$,

[7] We did robustness checks for the symmetry and constant multiplier assumptions. Specifically, we tested whether there might be asymmetric competition effects between Supercenters like Wal-Mart on other stores. We did not find significant differences. We also tested whether the distance-multipliers might be different for supercenters.

which is the global optimum of Eq. (10) where, $\overline{P}^*_{MLE}$ is the corresponding equilibrium CCPs, and Θ^*_{MLE} is the Maximum Likelihood Estimates (MLE).

Though the model imposes the restriction that the Conditional Choice Probabilities (CCPs), $\overline{P}^m$, for format and location entry should be in equilibrium (Eq. 10), we impose no additional structure to guarantee equilibrium uniqueness. Therefore, there could potentially be *multiple equilibria* (CCPs) *for the model* consistent with the true model parameters, Θ^0. Using a nested fixed-point (NFXP) approach for estimation is computationally demanding as it involves solving for the fixed-point of Eq. (6) at the trial value of Θ at each step of the likelihood maximization. More importantly, for a trial value of Θ, if Eq. (6) has multiple solutions for CCPs then the likelihood (Eq. 10) is not well defined. This is the problem of *multiple equilibria in the model*.[8] Hence, the NFXP approach is not well adapted to finding the MLE.

The Nested Pseudo Likelihood (NPL) approach developed by Aguirregabiria and Mira (2007) avoids the problem of multiple equilibria in the model by imposing the equilibrium condition for CCPs only for the final estimate of Θ and not for every trial value. The standard NPL approach starts with an initial guess of the CCPs, and converges to an equilibrium solution in the limit. For example, in our case, we would start with initial guess values for firms' beliefs about rivals' CCPs, $\overline{P}_0$. Then, using Eqs. (6) through (10) we would obtain the likelihood, $L(\overline{P}_0, \Theta)$. Maximizing the likelihood would give the parameter estimates, Θ_1. Using these parameter estimates in Eq. (6) would give new CCPs, $\overline{P}_1$, that is not necessarily an equilibrium associated with Θ_1. This would constitute one iteration and the new CCPs would be used as a guess for firms' beliefs about rivals' actions in the next iteration. The n^{th} iteration of the standard NPL approach can be denoted by the following contraction mapping:

$$\left(\overline{P}_n, \Theta_n\right) = M\left(\overline{P}_{n-1}\right) \;\; where, \;\; \Theta_n = \arg\max_{\Theta} L\left(\overline{P}_{n-1}, \Theta\right); \; \overline{P}_n = \Psi\left(\overline{P}_{n-1}, \Theta_n\right) \qquad (12)$$

With multiple iterations, if there is convergence, the contraction mapping would converge to an equilibrium solution or a *NPL fixed point*, $\left(\overline{P}^*, \Theta^*\right)$.

2.2.3 Obtaining the MLE

The MLE is the NPL fixed point that maximizes the likelihood function. However, the contraction mapping in Eq. (12) may not have a unique NPL fixed point, and the standard NPL approach may not necessarily return the MLE.[9] The multiple NPL fixed points are equilibrium solutions that are essentially the different 'local optima' of Eq. (10). Consequently, the NPL iterations may potentially converge to a 'local optima' and not the global optimum. Further, different starting values for $\overline{P}_0$ may lead to different 'local optima'. One option is to spread the search for the global optimum over a wide range of the contraction mapping, $M(\overline{P})$, by using *parallel-NPL*

[8] Note that the problem of *multiple equilibria in the data* is associated with only the true parameter set, Θ^0, whereas the problem of *multiple equilibria in the model* is associated with any trial value of the parameter set, Θ, within the NFXP estimation approach.

[9] Different NPL fixed points will give different likelihood values. Under the assumption of *single equilibrium in the data*, the interpretation of a NPL fixed point is that every observation in the data comes from that particular equilibrium.

where a large number of NPL algorithms, say, T, are run in parallel with different starting values. This approach, upon convergence, would give us a set of fixed points (many of which may be identical), $\left[\left(\overline{P}^{1*},\Theta^{1*}\right);\left(\overline{P}^{2*},\Theta^{2*}\right);\ldots;\left(\overline{P}^{T*},\Theta^{T*}\right)\right]$ However, it does not guarantee that this set will contain the global optimum, $\left(\overline{P}^{*}_{MLE},\Theta^{*}_{MLE}\right)$.

For a more efficient search of the global optimum, Aguirregabiria and Mira (2005) propose combining the parallel-NPL with a Genetic Algorithm (GA). GA is a search heuristic that mimics natural evolution processes such as '*selection*', '*crossover*' or '*reproduction*' and '*mutation*', and can be used to obtain the global optimum of complex optimization problems. Combining the parallel-NPL with GA has two advantages — (1) It spreads the search for the global optimum over a much wider range of the contraction mapping than what is feasible with just the parallel-NPL, and (2) The GA steers the tracks of the parallel-NPL iterations towards those regions of the contraction mapping that are more likely to contain the global optimum.[10]

2.2.4 Convergence

The algorithm may not converge to the global optimum if the contraction mapping does not have good local convergence properties around the global optimum. Kasahara and Shimotsu (2008) recommend transforming the mapping by replacing $\Psi\left(\overline{P},\Theta\right)$ with the following log-linear combination of $\Psi\left(\overline{P},\Theta\right)$ and $\overline{P}$:

$$\Lambda\left(\overline{P},\Theta\right)=\left[\Psi\left(\overline{P},\Theta\right)\right]^{\delta}\left[\overline{P}\right]^{1-\delta};\ \delta\in[0,1] \tag{13}$$

Note that $\overline{P}=\Lambda\left(\overline{P},\Theta\right)$ and $\overline{P}=\Psi\left(\overline{P},\Theta\right)$ have the same fixed-point solution(s). An appropriate value of δ can modify the concavity or convexity of the mapping such that the transformed mapping is *Locally Contractive* around the fixed point and will converge even if the original mapping does not.[11] Finally, even when the mapping does converge, the rate of convergence could be extremely slow and may require a large number of iterations. To avoid this, Kasahara and Shimotsu (2008) propose the following q-stage operator called *q-NPL*:

$$\Lambda^{q}\left(\overline{P},\Theta\right)=\underbrace{\Lambda\left(\Lambda\left(\ldots\left(\Lambda\left(\overline{P},\Theta\right),\Theta\right),\ldots,\Theta\right),\Theta\right)}_{q\ times} \tag{14}$$

[10] Su and Judd (2012) suggest using a Mathematical Programming with Equilibrium Constraints approach that finds the parameter estimates and the equilibrium CCPs simultaneously. However, like the parallel-NPL, this approach also relies on multiple runs with different starting values to find different equilibria. Hence, its ability to find the global optimum in problems that have a large action space (as in our entry and location choice problem) is unclear.

[11] Essentially, this means that the value of δ should be such that the eigen values of $\frac{\partial\Lambda\left(\overline{P},\Theta\right)}{\partial\overline{P}}$ are less than one. Kasahara and Shimotsu (2008) suggest the following procedure for selecting the value of δ: Simulate a sequence $\left\{\widetilde{P}_n\right\}_{n=0}^{N}$ by iterating the transformed mapping for different values of δ, say for δ {0.1, 0.2, …, 0.9}. Then pick the value of δ that leads to the smallest value of the mean of $\frac{\left\|\widetilde{P}_{n+1}-\widetilde{P}_n\right\|}{\left\|\widetilde{P}_n-\widetilde{P}_N\right\|}$ across $n=1,\ldots,N$.

Again, $\overline{P} = \Lambda^q\left(\overline{P}, \Theta\right)$ and $\overline{P} = \Psi\left(\overline{P}, \Theta\right)$ have the same fixed-point solution(s). In addition, $\Lambda^q\left(\overline{P}, \Theta\right)$ also has the locally contractive property of $\Lambda\left(\overline{P}, \Theta\right)$. Hence, in our estimation, we replace the standard NPL operator, Ψ, with the Locally Contractive, q-NPL operator, Λ^q. The resulting parallel NPL iterations are then combined with GA as described above. This procedure searches efficiently over the space of possible equilibria and converges fast to a set of equilibria which almost certainly contains the global optimum. To the best of our knowledge, this is the first empirical application of this procedure to speed convergence. Details of the sequence of steps involved in estimation are provided in Appendix A.

3 Data

3.1 Sample markets

An empirical analysis of firms' strategic entry, location and format decisions requires an appropriate set of *sample markets*. A market must be large enough to not only accommodate multiple competing firms of different formats, but also for the possibility of spatial differentiation among the entering firms. Yet, there is little value in studying spatial competition in too large a spatial market where firms would locate far away from each other (for grocery stores spatial competition falls off rapidly beyond 3 miles). In our application for big-box grocery retailers, extremely small towns and villages or large areas such as *Census Metropolitan Statistical Areas* are therefore not useful to be defined as a market. Further, towns and cities with very high population tend to have very complex zoning regulations such as sub-division zoning which cannot be inferred from our zoning data. Also, these large and dense markets usually have multiple stores of a retail chain requiring us to identify individual retailers and distinguish cannibalization effects from competition effects. Finally, clusters of towns, and suburbs of large cities like Chicago make it difficult to define a market boundary that reasonably separates retailers and consumers within a geographical area, where the market is self-contained, in being able to clearly define who are the consumers 'inside' a market and who are rivals 'outside' the market.

Given these challenges, we employ the following two criteria in selecting markets for analysis: (1) Single towns and town pairs with populations ranging from 20,000 to 250,000 people; (2) isolated markets that do not have another city or town within a 10 mile radius. Based on these criteria, we selected a set of 100 markets (i.e., M=100) across several U.S. states.[12] Figure 2 shows the spread of markets in our sample across the entire U.S. Table 1 lists descriptive statistics for our sample markets.

3.2 Consumers, grocery retailers, and their locations

Our empirical application corresponds to the year 2008 as we have grocery store location data for that year. Data on market characteristics are obtained from the U.S. Census. Although detailed demographic data at a Census Block Group (CBG) level

[12] Increasing the number of markets increases the computational burden, but with only marginal improvements in the precision of model parameters.

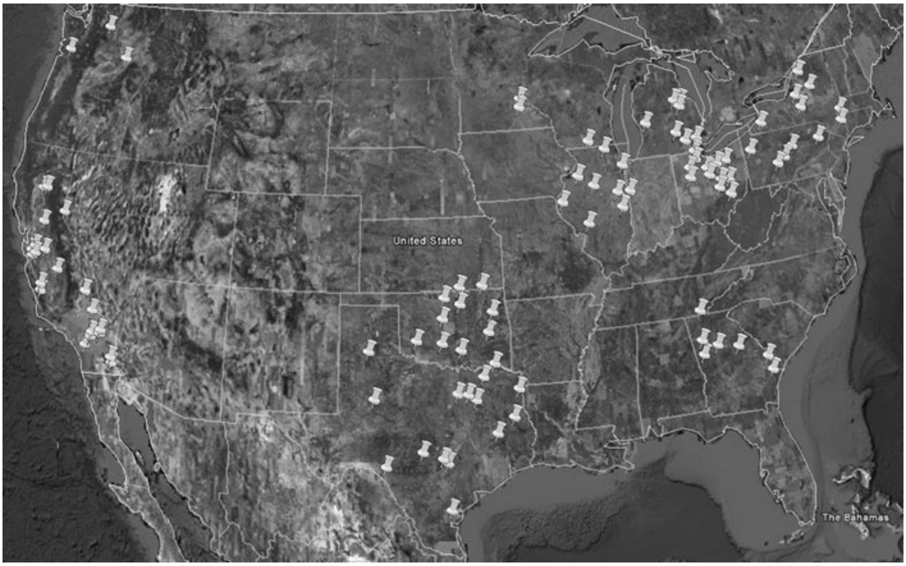


Fig. 2 Locations of 100 sample markets

are available only for the year 2000, the U.S. Census provides annual census projections for the county level. Hence, we project the CBG level census data to their 2008 values by the proportion of change in the respective counties between 2000 and 2008. As we do not have information about consumers beyond the CBG level, we follow the convention in the literature and place consumers in a CBG at the population-weighted center of the CBG. These are our *consumer locations*.

Geographic coordinates (latitude and longitude) of big-box grocery stores for 2008, and the store formats, are obtained from Nielsen's *'Trade Dimensions'*. In 2008, our 100 sample markets had altogether 751 big-box grocery stores. These stores have been classified into six format types (i.e., $F=6$): *Supermarkets* (SM), *Superstores* (SS), *Supercenters* and *Wholesale Clubs* (SC), *Limited Assortment* and *Warehouse stores* (LA), *Natural Foods* stores (NF) and *Food and Drug* stores (FD).[13] Table 2 provides a description of these store formats.

For the location choice game, we divide a market into a uniform grid of discrete 1 sq. mile blocks or *market locations*. Our 100 sample markets have a total of 7,216 such locations. But zoning regulations dictate which of these locations are available for big-box retailers. Below, we discuss our approach for identifying these *retail locations* and their *commercial centers*. Just as consumers are placed at the population-weighted center of CBGs, we place retailers within a retail location at the commercial center of the location.

Our concept of market locations differs from the standard approach in earlier research that treats census divisions as market locations and places retail stores at the population- weighted center along with consumers. The standard approach simplifies the data setup process but it has severe drawbacks: (1) The population-

[13] Many chains operate multiple formats (e.g., *Safeway, Vons Market, Target,* etc.), hence our assumption that chains can choose the optimal format for each location is reasonable. However, we note that a few chains operate only one format (e.g., *Whole Foods* operates only as a Natural Food store).

Table 1 Descriptive statistics of 100 sample markets

	Mean	Maximum	Minimum
Population (in 1000 s) scaled to 2007–08	84.8	249.7	20.7
Per capita income (in $ 1000 s) scaled to 2007–08	17.2	26.7	10.9
Area (in sq. miles) (= Number of discrete 1 sq. mile locations in a market)	73	225	16
Number of big-box grocery stores	7.5	21	2
Number of rival stores in 0–2 miles of a grocery store	1.3	9	0
Number of rival stores in 2–4 miles of a grocery store	2.8	12	1
Number of store formats	4	6	2
Proportion of market area zoned for retail stores	0.437	0.812	0.119

weighted center of a census division is likely to be a residential zone so that placing retail stores there would confound the inclusion of zoning regulations; (2) Stores are rarely present in the interior of a census division, rather, they are present on roads that border these census divisions; (3) Census divisions vary extensively in size so that, for large census divisions, stores may be located quite far from the center. In contrast, our approach allows us to incorporate spatial zoning, and it avoids major distortions of the distances between competitors and the distances of stores from population centers. Note that we use Great Circle distance as the distance between any two points.

We next describe the *National Land Cover Dataset* (NLCD) and discuss how it is used in conjunction with Geographical Information System tools such as *ArcGIS* and *Google Earth* to recover the potential retail locations and their commercial centers.

3.3 Spatial zoning data

Multi-Resolution Land Characteristics Consortium, a conglomerate of several federal agencies, has created two NLCD datasets that provide consistent and accurate digital land-cover information for the coterminous U.S. The first national land-cover mapping project, *NLCD 1992*, was derived from the early to mid-1990s *Landsat Thematic Mapper* satellite data. It applied a 21-class, geo-referenced, land-cover classification (see Vogelmann et al. 2001). The second project, *NLCD 2001*, updated the data for the year 2001 (see Homer et al. 2004). Both datasets have a spatial resolution of 30 m. That is, every 30 sq. meter area of land is classified as a specific land type (e.g., deciduous forest, grassland, open water, etc.) and is allocated one pixel point with a distinct color code and the associated latitude and longitude.[14] Interestingly, the land type classifications include residential and commercial land. Residential land is further classified into low and high intensity residential land, and commercial land comprises of highly developed areas that do not include residential areas. We use the NLCD data in the following three steps to identify the potential retail locations.

Step 1: *Constructing Market Boundaries and Market Locations:*

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

Table 2 Descriptive statistics of various grocery store formats

Store format	Supermarket (SM)	Superstore (SS)	Ltd. Assort. (LA)	Natural Food (NF)	Food+Drug (FD)	Supercenter (SC)
Examples of retailers[a]	Hi-Low Food Stores, Price Chopper, Vons Market	Jewel Food Store, BI-LO, Vons Market, Albertsons, Safeway,	Save-A-Lot, Price Rite, Aldi, Smart & Final	Whole Foods, Trader Joes	Jewel-Osco, Kroger, Albertsons, Safeway	Wal-Mart Supercenter, Super Target, Meijer, Sams Club, Costco
Total number of stores in 100 sample markets	164	106	127	39	189	126
Maximum number of stores in a market	8	4	6	3	7	5
Average store area (in sq. feet)	13,500	35,500	14,500	10,500	41,500	163,000
Average annual store revenue from grocery sales (in $ M)	5.93	15.24	5.23	9.22	16.09	51.84
Average ratio of grocery revenue to total store revenue	1	1	1	1	0.71	0.62

[a] Some retailers have more than one format (e.g., Vons, Albertsons, and Safeway). We follow the format classification of individual stores provided by AC Nielsen

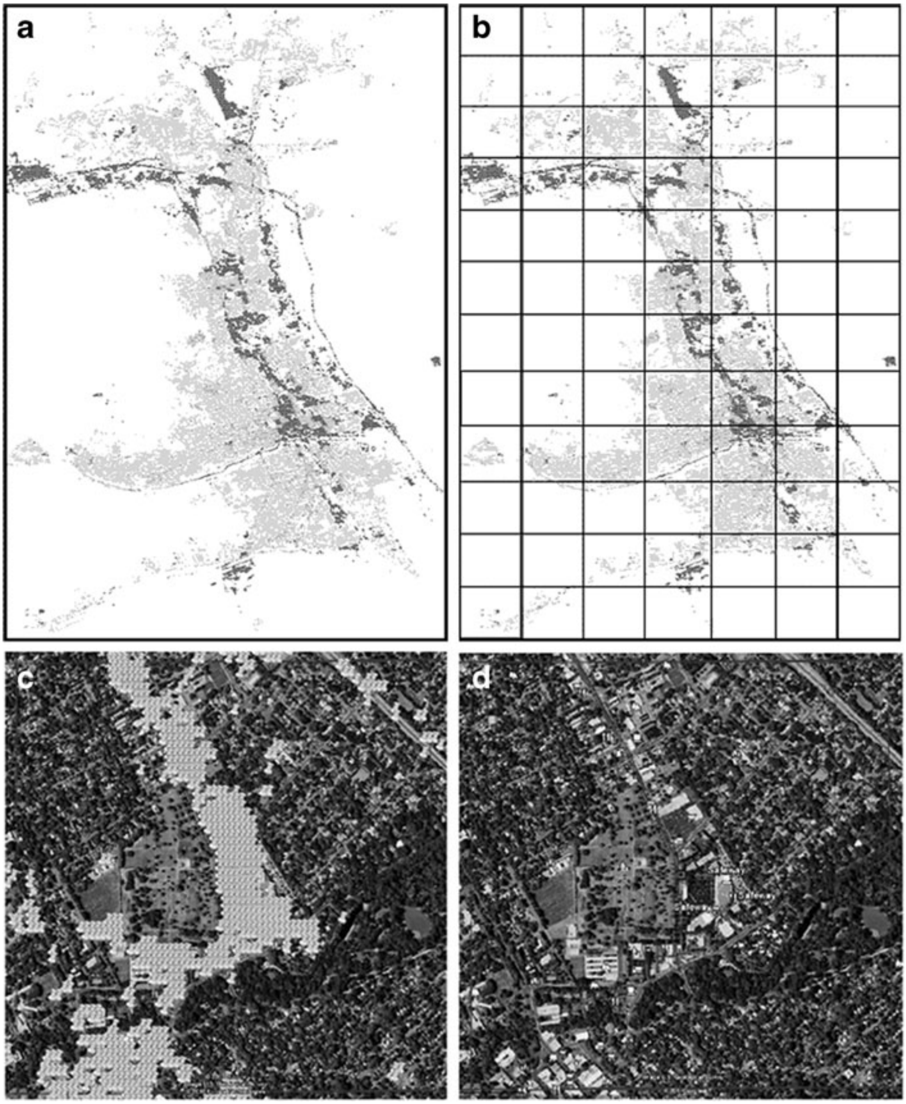


Fig. 3 **a** Constructing market boundaries based on visual inspection of residential and commercial pixel density. **b** Dividing a rectangular market into a grid of 1 sq. mile blocks or discrete locations. **c** Using commercial land pixel data to obtain extent of commercial activity within a location and the commercial center of the location. **d** Using '*Places of Interest*' in *Google Earth* to check for the presence of big-box stores in commercial locations

We use the data in *NLCD 2001* to construct the market boundaries of our sample markets. The residential and commercial land area pixel points in each market are projected on a map by using the *ArcGIS* software. This gives us the spatial area of interest for a market. A simple visual inspection of the pixel density is used to construct the market boundaries where the pixels fade away (See Fig. 3(a)). As our sample markets are reasonably isolated from other towns and cities, we can be flexible in choosing the shape of their boundaries. A rectangular shape is preferred so that a market can be easily divided into a uniform grid of discrete blocks or *market locations*. Thus, we construct imaginary rectangular borders (L miles X H miles where L and H are integers that vary across markets) around the residential and commercial pixel points of each market and then divide the market, specifically, into 1 sq. mile locations (See Fig. 3(b)).

Step 2: *Commercial Activity and Commercial Center in a Location:*

The extent of commercial activity in a location (as defined above) could affect firms' profit in the location if consumers have a preference for multi-purpose shopping or one-stop shopping. For instance, when shopping for groceries, consumers may like to combine their shopping trip with non-grocery purchases such as clothing and electronics so that locations with more retail businesses may be more attractive to firms. We isolate the *NLCD 2001* pixel points that correspond to commercial land with retail businesses (See Appendix B for more details) and use the number of pixel points in a location as a measure for the *extent of commercial activity* in that location. The mean of the latitudes and longitudes of the commercial land pixel points in a location gives us the *commercial center* of the location (See Fig. 3(c)). We place all retail stores within a location at the commercial center of that location. We prefer the commercial center to the geographical center for the placement of retail stores also because the commercial center is likely to coincide with unobservables such as the positions of major road intersections within the 1 square mile block locations.

Step 3: *Discerning Potential Retail Locations from other Commercial Locations:*

The market locations that contain the commercial land pixel points are the commercial locations and they constitute a very small share of all market locations. The locations without any commercial activity are mostly residential locations and some barren land. Hence, we account for residential zoning by excluding locations that do not have any commercial land pixel points. But even within commercial locations, not all locations may be open to big-box retailers. For instance, some commercial zones like, say, downtown areas, might only allow small businesses such as banks and restaurants. An obvious candidate for a potential *retail location* for big-box stores is any commercial location that has at least one big-box store— this could be either a grocery or non-grocery store. Hence, we project the locations on to *Google Earth* and use a tool called '*Places Categories*' which shows the locations of various types of businesses in a region (See Fig. 3(d)). We manually combed through the commercial locations, and specifically checked for the presence of major retail stores, major grocery stores and shopping centers to identify the commercial locations that have at least one big-box store.

Now, the absence of any type of big-box store in a commercial location does not necessarily imply that such stores are not allowed in that location. In particular, a commercial location that is open to big-box stores may not have any such store if it is in an unfavorable or poor neighborhood and cannot support a big store. As we do not have a precise method for identifying such locations, we use the following heuristic. For each market, we find the minimum value of the total income of consumers within a 2-mile radius of the commercial locations that have big-box stores.[15] We use this

[15] Our results are robust to using radii of 1, 2 and 3 miles.

minimum as a benchmark for a commercial location in the market to be attractive enough to support at least one big-box retail store. That is, if a commercial location does not have any big-box store and the total income of consumers within a 2-mile radius of the location is less than the market benchmark, then we presume that the absence of a big-box store is due to the unattractiveness of the location and not necessarily because of zoning restrictions. Hence, a commercial location with no big-box store is still treated as a potential retail location when the following condition is satisfied:

$$\begin{matrix} \textit{Income in 2-mile radius of a commercial} \\ \textit{location that has no big-box store} \end{matrix} \leq \left\{ \begin{matrix} \textit{Income in 2-mile radius of a commercial} \\ \textit{location that has a big-box store} \end{matrix} \right\}_{16}$$

To summarize, we use the NLCD data to construct market boundaries so that each market can be divided into a grid of 1 sq. mile locations. Then the commercial land pixel points are used to obtain the extent of commercial activity in a location and also to locate the commercial center of the location. Extant models that do not account for zoning, assume that firms are allowed to set up stores in any market location. In contrast, we account for residential zoning by excluding locations that do not have any commercial land pixel points. Finally, we account for zoning regulations particularly against big-box retailers, within commercial locations, by defining potential retail locations as those commercial locations that (1) have at least one big-box store which is either a grocery or a non-grocery store, and (2) do not have a big-box store and are in a poor neighborhood which is below the market benchmark as described above.

4 Results

We begin with a descriptive analysis of the data before reporting the results of the structural model estimation.

4.1 Descriptive analysis

The descriptive analysis has two objectives. First, we test for empirical evidence in support of the theoretical conjectures about how tightened spatial restrictions impact retail entry and format variety without any model restrictions. Second, we estimate the correlations between zoning restrictions and factors affecting market potential and competitive intensity to obtain guidance on the expected magnitude and direction of the potential bias from omitting zoning restrictions.

[16] This procedure may exclude some locations in high-income neighborhoods that are above the income benchmark, but are actually available for big-box retailers. To the extent, our criterion checks for any type of big-box store (grocery or non-grocery stores), such errors are likely infrequent, with little impact on estimates. The procedure may also include locations in low-income neighborhoods that are below the income benchmark even though in reality big-box stores are not allowed in such locations. Such erroneous inclusion of low population or low per capita income locations would cause overestimation of the positive profit impact of these variables.

4.1.1 The link between spatial and format differentiation

We report the relationship between zoning restrictions and the number of competing stores in the 5 mi. radius of a given big-box grocery store in Table 3(a).[17] We operationalize zoning restrictiveness with two variables: (1) proportion of area that is available for entry in the 5 mile radius of a store (*Proportion of Available Area*); and (2) proportion of that available area which is concentrated within a 3 mile radius of the store (*Proportion within 3 mi.*) — ceteris paribus, a higher value of *Proportion within 3 mi.* indicates a spatially tighter zoning around the central location or a lesser scope for spatial differentiation.[18] Controlling for population and per-capita income, we find an insignificant relationship for *Proportion Available Area*, but *Proportion within 3 mi.* has a negative and significant impact, consistent with the theory that tightened zoning restrictions significantly reduce entry.

We report the relationship between zoning restrictions on format concentration (the opposite of format variety) in Table 3(b). We operationalize format concentration through a metric that is similar to the Herfindahl-Hirschman Index (HHI); i.e., we use the sum of squares of shares of the six different grocery store formats. Hence, a higher value for format HHI would imply a greater format concentration and less format variety. We find a negative and significant relationship between format concentration and *Proportion Available Area*; i.e., loose zoning restrictions allow a large variety of stores to enter. Further, there is a negative and significant relationship between format concentration and *Proportion within 3 mi.*; showing that controlling for the available area, format variety increases with tighter zoning restrictions as predicted by the theory.

Collectively, these results demonstrate that zoning restrictions do negatively impact retail entry and format concentration. Further, given an available area for entry, and controlling for the number of entrants, tightening zoning restrictions that prevent firms to spatially differentiate (i.e., *Proportion within 3 mi.*), causes greater format differentiation and variety. Overall, the results support the theory that firms will respond to the inability to spatially differentiate through product (format) differentiation. Given this descriptive evidence, we will further explore the predictive implications of zoning restrictions for entry and format variety through counterfactual simulations based on estimates from a structural model of entry, location and format choice.

4.1.2 Potential biases due to omission of zoning

We next discuss our expectations about the type of biases due omission of zoning restrictions. We report the histogram of the point-biserial correlation between zoning restrictions and demographic variables such as population and per-capita income for the 100 sample markets in Fig. 4. The figure indicates that there is substantial heterogeneity in the correlations. In the aggregate, the mean correlation with

[17] Popular business press articles suggest that the trade radius of a Supermarket is typically about 2–3 miles and for a Supercenter like Wal-Mart supercenter it is about 5–7 miles. We tested with different trading radii around a given store and obtained similar results.

[18] Let the total area within the 5 miles radius of store be approximated as 75 square miles (approximation of $\pi * 5^2$). Suppose 25 square miles of this area are available, then *Proportion Available Area*=0.3. Suppose 20 square miles of this available area is concentrated within the 3 mile radius, then *Proportion within 3 mile*=0.8.

Table 3 OLS regressions (preliminary analyses) for impact of zoning on (a) number of competing stores, and (b) format concentration, within 5 mi. of a grocery store

Variable	3(a): Dependent Variable= Number of Competing Big-Box Grocery Stores	3(b): Dependent Variable= HHI of Grocery Store Format Concentration
Intercept	0.172	**0.792*** **
Population (10,000 s)	**1.312*** **	−0.007*
Per-capita income ($10,000)	**1.198*** **	**−0.042*** **
Proportion available area	0.307	**−0.132*** **
Proportion within 3 mi.	**−2.035*** **	**−0.063** **
Number of rivals	–	**−0.068*** **
Square of number of rivals	–	**0.003*** **
R-Sq	**79.4 %**	**50.8 %**

All significant estimates in bold

*p<0.1, **p<0.05, ***p<0.01

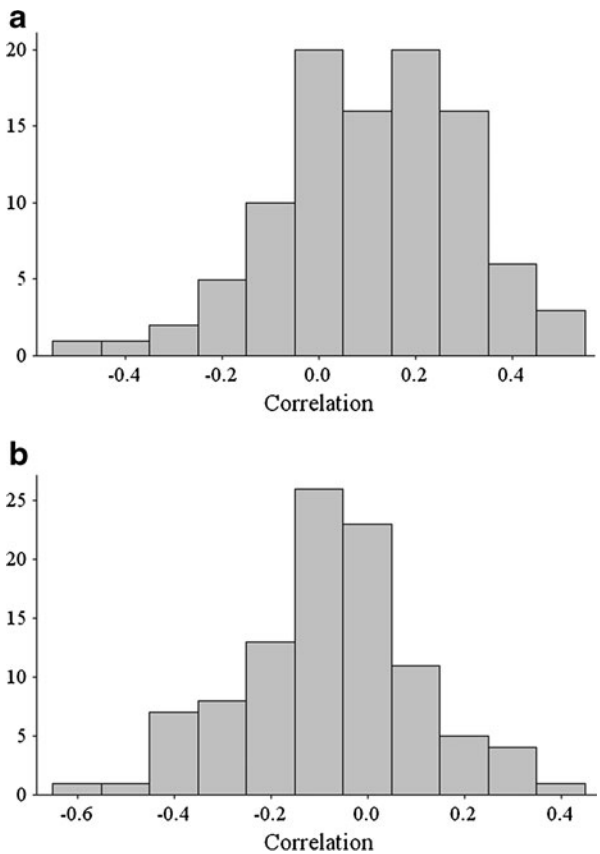


Fig. 4 **a** Histogram of (market-level) point-biserial correlations between locations available for entry and *distance-weighted population* within 5 mi. of the location. **b** Histogram of (market-level) point-biserial correlations between locations available for entry and *distance-weighted per-capita income* within 5 mi. of the location

distance-weighted population in a 5 mile radius is 0.1219 (Std. deviation=0.1993) and mean correlation with distance-weighted PCI is −0.0889 (Std. deviation= 0.1849). Thus on average, commercial zoned locations are slightly positively (negatively) correlated with distance-weighted population (distance-weighted PCI). Hence, when zoning is omitted, the presence of retailers closer to (farther from) locations with high population (PCI) may be misattributed to a stronger (weaker) effect of population (PCI) on store profits. That is, a model that omits zoning is likely to overestimate (underestimate) the effect of population (PCI). Nevertheless, as the standard deviations indicate, there is lot of variation in the extent of correlation in different markets; this suggests that the bias at the level of each market would be different, and only accommodating actual zoning restrictions at the market level would help evaluate the correct counterfactuals.

The impact of omitted zoning on the estimates of spatial competition parameters is harder to predict. Consider the following stylized specification for firms' profit at a location

$$\pi = \beta_0 + \beta_1 X + \beta_2 N + \beta_3 Z + \varepsilon$$

where X includes market characteristics like population and income, N is the number of competitors at the location, and, for the purpose of this illustration, Z is an indicator for zoning such that Z=0 if zoning restricts entry into the location and Z= 1 otherwise. The competition parameter, β_2, will be negative, and β_3 will be positive. If zoning is omitted, we would estimate the mis-specified model

$$\pi = \beta_0 + \beta_1 X + \beta_2 N + \varepsilon^*$$

where the error term, $\varepsilon^* = \beta_3 Z + \varepsilon$, includes the omitted zoning variable. Then the bias in the estimate of the competition parameter will have the same sign as $(\rho_{NZ} - \rho_{XZ} * \rho_{XN})$ where ρ_{ij} is the correlation between variables i and j.[19] As competitors do not exist in locations where entry is restricted, we have $\rho_{NZ}>0$. Now, if zoning restrictions are not related to market characteristics then $\rho_{XZ}=0$, and the bias is positive. That is, in this case, the competition effect will be underestimated. However, if zoning restricts entry into locations with higher population and income then $\rho_{XZ}<0$. Also, due to the absence of competitors in such locations, we may find that $\rho_{XN}<0$. Then, depending on the relative magnitudes of ρ_{NZ} and $\rho_{XZ}*\rho_{XN}$, the bias can be positive (i.e., competition effect will be underestimated) or negative (i.e., competition effect will be overestimated). Hence, the actual bias on competition effects will depend on how zoning is related to the incomes and populations in the market.

4.2 Structural model estimates

Our search for the MLE by combining parallel-NPL with GA led to a set of identical NPL fixed points, which suggests that for our empirical application the model has a unique equilibrium solution or a unique NPL fixed point, which is also the MLE. We first discuss the results with homogeneous retailers where we do not distinguish

[19] Under the misspecified model, the expected value for β_2 is given by Hanushek and Jackson (1977) as $E[\beta_2] = \beta_2 + \beta_3 b$ where the bias $b = \frac{(\rho_{NZ} - \rho_{XZ} * \rho_{XN})}{1 - \rho_{XN}^2} * \frac{\sigma_Z}{\sigma_N}$, ρ_{ij} is the correlation between variables i and j, and σ_i is standard deviation of variable i.

between different grocery store formats similar to Seim (2006), before considering the model with heterogeneous retailers that can differentiate on formats.

4.2.1 Homogeneous retailers

In this model, we ignore firms' store format choice and assume that they only make entry and location choice decisions. This case is similar to the application in Seim (2006). Recall that the area around each location is divided into five distance bands of 1 mile widths (0–1 mi., 1–2 mi, 2–3 mi., 3–4 mi. and 4–5 mi.). For the observable market characteristics that affect store profit in a location we use population (*Pop*) and per capita income (*PCI*) of consumers in the different distance bands around the location, and the extent of commercialization (*Commercialization*) at the location. Table 4 presents the parameter estimates with zoning restrictions (Column 1) and without zoning restrictions (Column 2), where retailers can set up stores anywhere in a market. In both cases, the potential number of entrants (R) in each market is fixed at 25. Selecting a different value for R only alters the distribution of the unobserved market attractiveness parameter, ξ^m, but with negligible difference in the substantive results.

Not accounting for zoning generally gives downward biased estimates. Without zoning, commercialization has a higher coefficient because firm choice of location is

Table 4 *Homogeneous retailers* with and without controlling for zoning

Variable		Column 1 (Estimates *with Zoning*; # Retail Locations=2853)	Column 2 (Estimates *without Zoning*; # Retail Locations=7216)
Commercialization		**0.103*** **	**0.199*** **
Population coefficients	0–1 mi.	**1.167*** **	**1.626*** **
	1–2 mi.	**0.845*** **	**0.811*** **
	2–3 mi.	**0.322*** **	0.069
	3–4 mi.	**0.272*** **	−0.037
	4–5 mi.	−0.267	**−0.482*** **
Per capita income coefficients	0–1 mi.	**0.766*** **	**0.623*** **
	1–2 mi.	**0.268*** **	0.089
	2–3 mi.	−0.059	−0.042
	3–4 mi.	−0.114	−0.086
	4–5 mi.	−0.126*	−0.124*
Competition effect coefficients	0–1 mi.	**−0.615*** **	**−0.878*** **
	1–2 mi.	**−0.221*** **	**−0.139** **
	2–3 mi.	−0.084	0.053
	3–4 mi.	−0.021	0.098
	4–5 mi.	−0.055	0.068
μ (ξ)		**−6.603*** **	**−7.283*** **
σ (ξ)		**0.788*** **	**0.681*** **

All significant estimates in bold

*p<0.1, **p<0.05, ***p<0.01

attributed to the value of commercialization, rather than the fact that one can locate the store only in these zoned areas. As discussed above, available locations in our sample markets are positively correlated with population and slightly negatively correlated with per capita income (PCI). Hence, as expected, when zoning is omitted, the model overestimates the profit impact of population (1.626 without zoning versus 1.167 with zoning in the 0–1 mile band) and slightly underestimates the impact of PCI (0.623 without zoning versus 0.766 with zoning in the 0–1 mile band). Nevertheless, in both cases the profit impact of population and PCI decrease gradually with distance.

Like the impact of market characteristics, the spatial competition effect between rivals also decreases dramatically with distance. This signifies the benefit of spatial differentiation. In terms of bias, as expected the effects are mixed. We find that the competition effect at short distances within 0–1 mile is overestimated (−0.878 without zoning versus −0.615 with zoning) when we omit zoning. In contrast, the competition effect at moderate distances of 1–2 miles, say, is underestimated (−0.139 without zoning versus −0.221 with zoning). Finally, when we do not control for zoning, low entry into markets with more restrictive zoning is explained away by a low value for the unobserved market fixed effect (Mean value of ξ^m is −7.2827 without zoning versus −6.603 with zoning).

4.3 Inter-format and intra-format competition

We next consider heterogeneous retailers where firms make endogenous entry, location and store format choices. We classify grocery stores into six format types (i.e., F=6): *Supermarkets* (**SM**), *Superstores* (**SS**), *Limited Assortment* and *Warehouse stores* (**LA**), *Natural Foods* stores (**NF**), *Food and Drug* stores (**FD**) and *Supercenters* and *Wholesale Clubs* (**SC**). We again fix the potential number of entrants (R) at 25. Tables 5 and 6 present the parameter estimates for the cases with and without zoning, respectively.[20]

We can see that the impact of the observable market characteristics is different for different store formats. For instance, Table 5 shows that unlike the other formats, *Supercenters* are not attracted towards locations with high population rather they are more sensitive to population at farther distances of 2–4 miles, perhaps because of the high cost of operating this large format in densely populated areas and also because *Supercenters* are likely to draw shoppers from farther distances. Similarly, consumers' per capita income has a much greater positive impact on the store profit of *Superstores* than that of *Supermarkets* and *Supercenters*. Hence, the attractiveness of locations within a market varies across store formats. A comparison of the estimates again shows that ignoring zoning overestimates the impact of population and underestimates the impact of PCI.

The spatial competition parameter estimates in Table 5 reveal some interesting insights. First, in general, intra-format competition is greater than inter-format competition effects. A notable exception is that *Supermarkets* compete more intensely

[20] Based on results from preliminary estimations, we estimate separate coefficients for population at different distances for the *Supercenter* format, but common population-distance multipliers (Eq. (11.1)) for the other formats. However PCI-distance multipliers were not different for different formats; hence we report results with common multipliers for PCI.

Table 5 Store formats with *controlling for zoning restrictions*

Variable		Supermarket (SM)	Superstore (SS)	Ltd. Assort. (LA)	Natural Food (NF)	Food & Drug (FD)	Supercenter (SC)
Constant		na	**−2. 013*****	**−1.937****	**−3.193*****	**−0.939*****	−0.339
Commercialization		**0.078*****	**0.095*****	**0.106*****	**0.162*****	**0.124*****	**0.106*****
Population	0–1 mi. coefficient	**1.757*****	**1.628*****	**1.308*****	**2.028*****	**1.914*****	−0.245
	1–2 mi. multiplier/coeff.			**0.704*****			**0.872*****
	2–3 mi. multiplier/coeff.			**0.270*****			**1.051*****
	3–4 mi. multiplier/ coeff.			**0.258*****			**0.586*****
	4–5 mi. multiplier/ coeff.			−0.087			−0.049
Per capita income	0–1 mi. coefficient	**0.368****	**1.191*****	**0.928*****	**0.838****	**0.940*****	**0.718*****
	1–2 mi. multiplier				**0.327****		
	2–3 mi. multiplier				−0.079		
	3–4 mi. multiplier				−0.201		
	4–5 mi. multiplier				−0.212		
Competition effect	**SM**; 0–1 mi. coeff.	**−1.842****					
	SS; 0–1 mi. coeff.	**0.444****	**−2.300****				
	LA; 0–1 mi. coeff.	0.353*	**−2.481****	**−2.727*****			
	NF; 0–1 mi. coeff.	0.025	−2.403	−0.791	−4.417		
	FD; 0–1 mi. coeff.	−1.405	−0.170	**0.790****	−2.257	**−2.141****	
	SC; 0–1 mi. coeff.	**−2.188*****	0.810	**1.644****	2.088	−1.128	**−4.572*****
	1–2 mi. multiplier				**0.486*****		
	2–3 mi. multiplier				**0.278*****		
	3–4 mi. multiplier				**0.172*****		
	4–5 mi. multiplier				0.238		
μ (ξ)					**−7.297*****		
σ (ξ)					**0.998*****		

For Supercenter format, estimates corresponding to population at different distance bands are coefficients rather than distance multipliers

All significant estimates in bold

*p<0.1, **p<0.05, ***p<0.01

Table 6 Store formats without controlling for zoning restrictions

Variable		Supermarket (SM)	Superstore (SS)	Ltd. Assort. (LA)	Natural Food (NF)	Food+Drug (FD)	Supercenter (SC)
Constant		na	**−1. 311*****	**−1.272*****	**−2.583*****	**−0.812*****	−0.341
Commercialization		**0.171*****	**0.195*****	**0.205*****	**0.248*****	**0.221*****	**0.200*****
Population	0–1 mi. coefficient	**2.048*****	**2.193*****	**2.020*****	**2.525*****	**2.359*****	0.138
	1–2 mi. multiplier/coeff.			**0.555*****			**0.596****
	2–3 mi. multiplier/coeff.			0.084			**0.814*****
	3–4 mi. multiplier/ coeff.			0.066			0.301*
	4–5 mi. multiplier/ coeff.			−0.174			−0.185
Per capita income	0–1 mi. coefficient	**0.277*****	**0.907*****	**0.812*****	**0.688*****	**0.814*****	**0.525*****
	1–2 mi. multiplier			**0.144****			
	2–3 mi. multiplier			−0.067			
	3–4 mi. multiplier			−0.201			
	4–5 mi. multiplier			−0.241			
Compe-tition effect	**SM**; 0–1 mi. coeff.	**−1.753*****					
	SS; 0–1 mi. coeff.	−0.872*	**−0.506*****				
	LA; 0–1 mi. coeff.	−0.422*	−1.504*	**−1.389*****			
	NF; 0–1 mi. coeff.	−0.138	−2.186	−1.930	−1.938		
	FD; 0–1 mi. coeff.	−0.567	−1.769*	−1.358*	−3.047	**−2.131*****	
	SC; 0–1 mi. coeff.	**−2.549*****	−0.225	**0.952****	1.488	0.512	**−4.788*****
	1–2 mi. multiplier			**0.292*****			
	2–3 mi. multiplier			0.075			
	3–4 mi. multiplier			−0.051			
	4–5 mi. multiplier			0.077			
μ (ξ)				**−8.311*****			
σ (ξ)				**0.875*****			

All significant estimates in bold

*p<0.1, **p<0.05, ***p<0.01

with *Supercenters* (competition effect at 0–1 mi.=−2.188) than with other *Supermarkets* (competition effect at 0–1 mi.=−1.842). This effect is consistent with the conventional wisdom that the arrival of larger stores may have significant asymmetric impact on smaller stores. Finally, some inter-format effects are positive. This effect is consistent with the conventional wisdom that consumers may wish to shop at stores of different formats: they may shop mostly at a supercenter and then shop at a close by natural food store or limited assortment store for select products—suggesting agglomeration benefits. [21] Some of the inter-format competition parameters are even positive which indicates that certain store formats gain some agglomeration benefits when they locate close to a rival with a different format. For example, the *Limited Assortment* stores benefit from locating close to *Supercenters* (1.644) and *Food and Drug* stores (0.790). The multiplier parameters for the different distance bands again indicate that the competition effect decreases with distance from rivals.

Again omission of zoning leads to bias in competition parameters, with the direction of the bias mixed. Comparing with Table 6, we again see that omission of zoning underestimates the competition effect between rivals at moderate distances of 1–4 miles. However, several of the inter-format competition effects at short distances (0–1 mile) get overestimated, while most of the intra-format competition effects and the relatively high negative inter-format competition effects are underestimated.

5 Counterfactual simulations

We perform two sets of counterfactuals. First we assess how constraining the area available for retail entry through zoning affects retail entry and format variety. Second, we assess how certain prototypical zoning arrangements such as centralized zoning, neighborhood zoning and outskirt zoning impact format variety.

5.1 Effect of zoning restrictions on entry and format variety

To investigate the impact of greater zoning restrictions on firms' entry decisions and format variety, we generate 100, 8 mi. × 8 mi., hypothetical markets each of which is divided into 64 1 sq. mile block locations. Values for Population, PCI and commercial activity are randomly assigned to each market location. Starting with no zoning restrictions (i.e., all 64 locations in each market are available for retailers), we gradually increase zoning restrictions in these markets and explore the influence of zoning on market structure. For this, we reduce the number of locations that are available for retailers in a market in steps of one location. At each step, the new location that is 'zoned out' is randomly selected in each market so that at any step the scope for spatial differentiation varies across markets.

As we gradually increase zoning restrictions, we calculate firms' entry probability in each market and the equilibrium location and format choice probabilities. Note that given model parameter estimates, and the equilibrium number of firms that enter a

[21] An alternative explanation is unobservable to researcher location shocks (e.g., favorable road intersections or low rents) that may be observable to retailers. We thank a reviewer for suggesting this alternative explanation.

market, the fixed point solution of Eq. 6 will give us the equilibrium conditional location and format choice probabilities. We use the nested fixed point approach for this calculation.[22] Now, to obtain the equilibrium number of firms that enter a market, we calculate the entry probability using Eq. 7. For this, we need the unobserved market-specific components, ξ^m. We fix ξ^m for our simulated markets by using Eq. 9 and assuming that when there are no zoning restrictions ten out of 25 potential entrants ($N^m=10$; $R=25$) can enter each market.

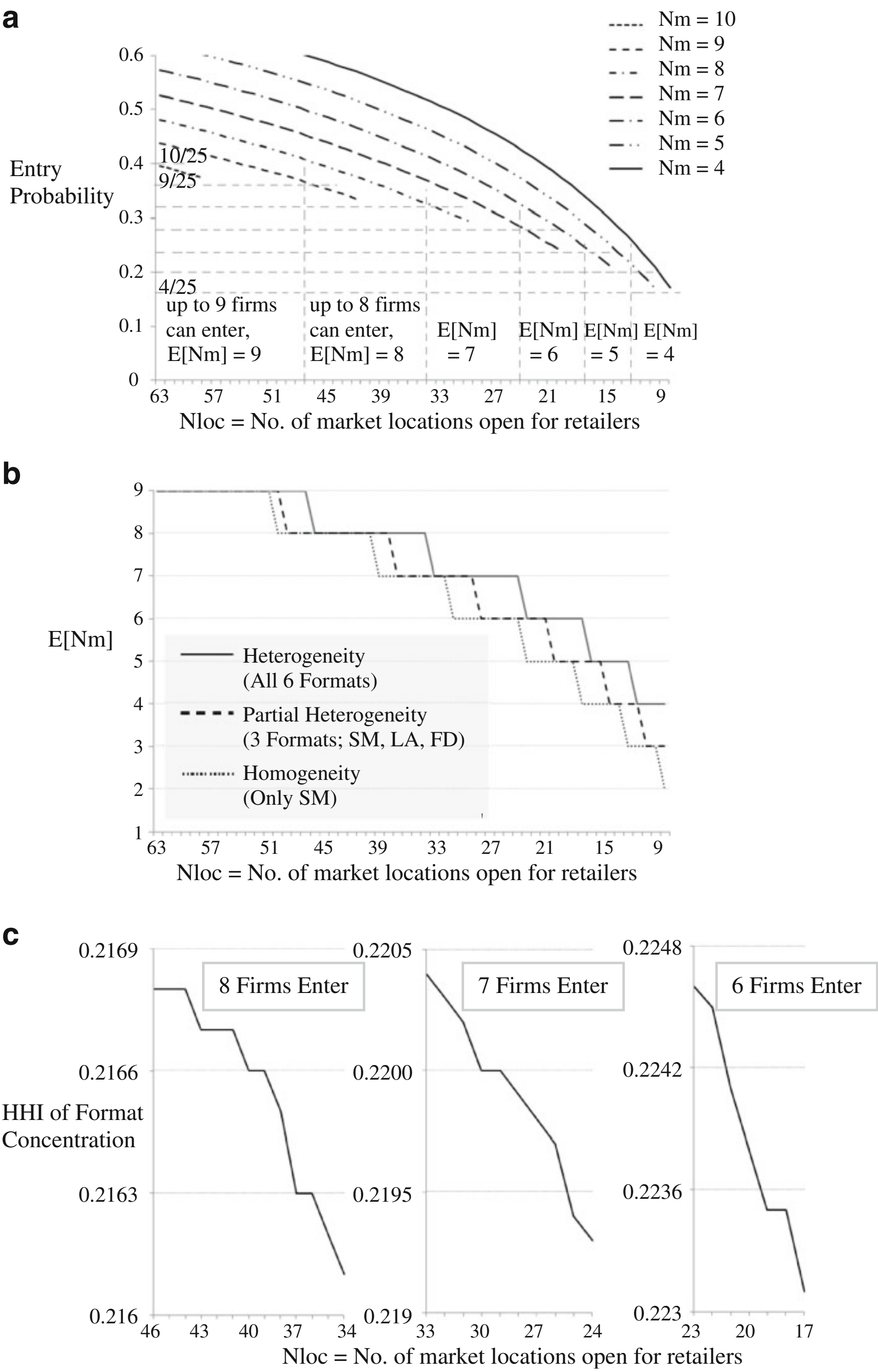
Figure 5(a) plots the average entry probabilities for our simulated markets when markets are heterogeneous. For this we use the estimates in Table 5. We see that entry probabilities decrease when fewer market locations are available to retailers. Consequently, the equilibrium number of retailers that can enter these markets decreases rapidly from ten firms when all 64 locations are available to five firms when only 16 locations are available; i.e., when the share of commercial zoned locations is only 25 % of the market area, the number of firms fall by 50 %.

We next explore how format differentiation can mitigate the entry deterring effect of zoning. For this we compare the entry probabilities for heterogeneous firms (all six store formats) with that under the assumptions of (a) partial heterogeneity where an entering firm can choose from one of three formats — *Supermarket, Limited Assortment* and *Food and Drug*, and (b) homogeneity where an entering firm can only setup a store with the *Supermarket* format. Now when zoning restrictions increase in markets, the entry probabilities drop at a faster rate under partial heterogeneity as the scope for spatial differentiation as well as format differentiation are limited. The drop in entry probabilities is even faster under homogeneity. This is shown in Fig. 5 (b) where we see that a greater heterogeneity on the format dimension increases the entering firms' ability to withstand greater zoning restrictions. For instance, in the case of heterogeneous firms, as we reduce the number of available market locations from 46 to 34, the average number of market entrants continues to remain fixed at eight firms. But in the case of homogeneous firms, the number of firms dropped to seven when the number of available locations was reduced to 39. Further, when the number of available locations falls to 8, i.e., 12.5 % of the available area, the number of entrants under heterogeneity (four entrants) is 50 % more than when there is only partial heterogeneity (three entrants) and 100 % more than when there is homogeneity (two entrants).

We next evaluate how the format mix or variety changes with greater zoning restrictions. We measure format concentration in a market through a metric that is similar to the Herfindahl-Hirschman Index (HHI). Specifically, we use the sum of squares of shares of the six different store formats in a market to obtain the following HHI of format concentration for a given number of entering firms:

$$HHI(N^m) = \sum_{f=1}^{F} \left[\sum_{l=1}^{l_m} p_{fl}(N^m) \right]^2 \tag{15}$$

[22] To deal with potential multiple equilibria in the model, we solved for the nested fixed point(s) for each market by starting with 1000 different guess values for the matrix of location and format choice probabilities. In roughly 95 % of the cases, we obtain unique equilibria for all guess values. For the remaining cases where they converge to different equilibria, we take the average choice probabilities over the 1000 fixed point solutions.


a
Nm = 10
Nm = 9
Nm = 8
Nm = 7
Nm = 6
Nm = 5
Nm = 4
Entry Probability
0.6
0.5
0.4
0.3
0.2
0.1
0
10/25
9/25
4/25
up to 9 firms can enter, E[Nm] = 9
up to 8 firms can enter, E[Nm] = 8
E[Nm] = 7
E[Nm] = 6
E[Nm] = 5
E[Nm] = 4
63 57 51 45 39 33 27 21 15 9
Nloc = No. of market locations open for retailers
b
E[Nm]
9
8
7
6
5
4
3
2
1
Heterogeneity (All 6 Formats)
Partial Heterogeneity (3 Formats; SM, LA, FD)
Homogeneity (Only SM)
63 57 51 45 39 33 27 21 15 9
Nloc = No. of market locations open for retailers
c
HHI of Format Concentration
0.2169
0.2166
0.2163
0.216
8 Firms Enter
46 43 40 37 34
0.2205
0.2200
0.2195
0.219
7 Firms Enter
33 30 27 24
0.2248
0.2242
0.2236
0.223
6 Firms Enter
23 20 17
Nloc = No. of market locations open for retailers

◀ **Fig. 5** **a** Equilibrium number of entrants: E[Nm]=k iff k/25≤Entry Prob(Nm=k) and (k+1)/25>Entry Prob(Nm=k+1). As zoning restrictions increase, the equilibrium number of heterogeneous firms that enter a market (E[Nm]) decreases. **b** As zoning restrictions increase, the equilibrium number of entrants decreases faster when store formats are more homogeneous. **c** Firms resort to greater format differentiation (lower *HHI*) under more restrictive zoning such that the equilibrium number of entrants may not change for large ranges of zoning restrictions. **d** Counterfactual exercise for three prototypical zoning patterns. In each case, 16 (out of 64) locations are available for retailers, and five retailers are assumed to enter a market. **e** Share of store formats and HHI of format concentration under different prototypical zoning patterns

A lower value for this *HHI* indicates a lower concentration of any single store format in the market or a greater format variety in the market. As shown in Fig. 5(c), we find that as zoning restrictions in markets increase, for large ranges of zoning restrictions the average HHI of format concentration decreases while the number of entrants remains fixed. For instance, as the number of available locations in the 8 mi.×8 mi. markets drops from 32 (50 % of the market) to 24 (37.5 % of the market), the HHI of format concentration decreases from 0.2203 to 0.2193 while the number of entering firms remains unchanged at seven. Hence, firms resort to format differentiation to mitigate the entry deterrent effect of zoning. Our results demonstrate that citizens and local governments making zoning decisions to restrict the available locations for retailers will not only reduce the number of retailers available to shop, but can also impact the type of shopping formats available to the citizenry.

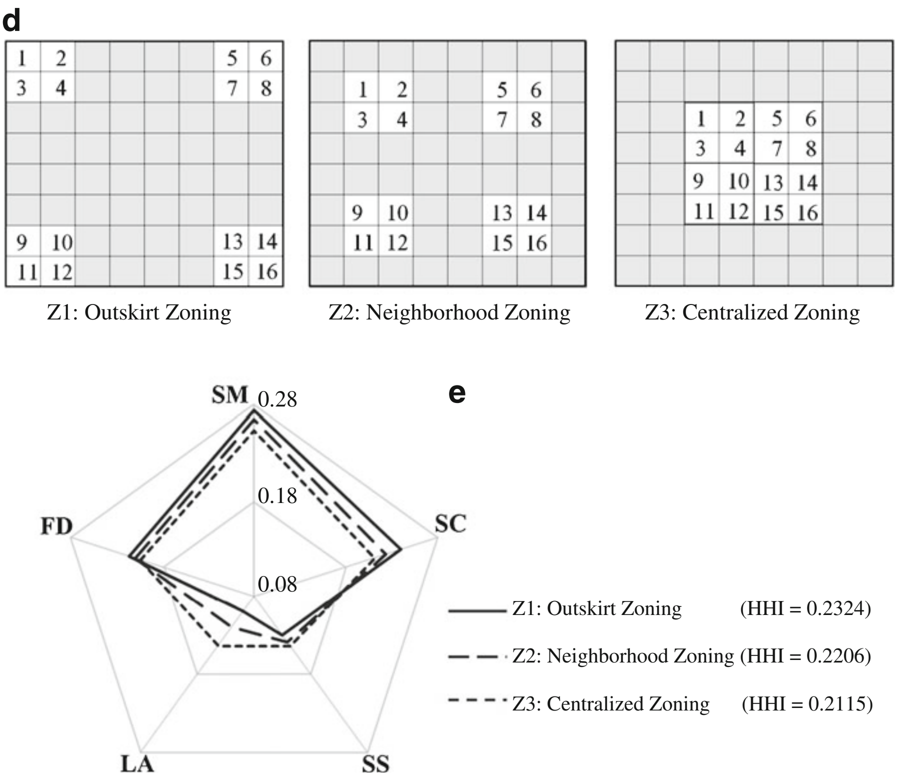


Fig. 5 (continued)

5.2 Prototypical zoning arrangements: centralized, neighborhood and outskirt zoning

We consider three prototypical zoning arrangements and how it impacts format variety: "centralized," "neighborhood," and "outskirt" zoning. Centralized zoning seeks to mimic a zoning arrangement where commercial locations are restricted to a town center; neighborhood zoning seeks to mimic an arrangement where commercial locations are restricted to the centers of local neighborhoods, while outskirt zoning seeks to mimic a zoning arrangement where commercial locations are restricted to the periphery of the town. These arrangements differ in the extent to which firms can spatially differentiate.

As before we take 8 mi.×8 mi. markets with 64 1 sq. mile locations, but create zoning restrictions consistent with the three prototypical zoning arrangements. To control for availability of locations across the three zoning arrangements, we allow entry into only 16 of the available 64 locations in all arrangements. Those 16 locations are distributed in the three arrangements as shown in Fig. 5(d). The 16 locations are divided into four 2 mi×2 mi. retail zones that are at the periphery (Z1: Outskirt Zoning) or at the center of neighborhoods (Z2: Neighborhood Zoning) or at the center of the market (Z3: Centralized Zoning). Similar to our previous counter-factual, we assume that five firms enter a market, irrespective of the zoning pattern.

Figure 5(e) shows that as the zoning pattern changes from *outskirt* to *centralized*, the more popular, *Supermarket*, *Supercenter* and *Food and Drug* store formats lose market share whereas *Limited Assortment* and the *Superstore* formats gain share with centralized zoning.[23] Effectively, the HHI of format concentration decreases from 0.2324 under outskirt zoning to 0.2115 under centralized zoning.

Overall, we conclude that centralized zoning leads to greater format variety with a number of smaller formats, relative to outskirt zoning which leads to more homogeneous larger format zoning. Our results should inform recent debates about the homogenization of retail formats as towns open up peripheral locations for development of big-box retailers.

6 Conclusion

The literature on retailer entry and location choices has thus far ignored the spatial zoning regulations that impact entry and location decisions. Taking advantage of a publicly available, digital land cover database, NLCD, we are able to study the effect of zoning on entry, location and format choices. We estimate a static, structural, simultaneous move game model of entry, location and format choice with incomplete information using data on the observed choices of big-box grocery store retailers in a national sample of markets. We use recent advances in the empirical estimation literature of discrete games to address issues of multiple equilibria in the model and data as well as problems due to slow convergence of the estimation algorithm.

Our analysis leads to the following key takeaways: First, zoning reduces entry because the inability to spatially differentiate increases competition and reduces profitability and the number of firms a market can support. However, since inter-format competition is much less intense than intra-format competition, firms resort to

[23] The *Natural Food* format also gains a small amount of market share but it is not shown here for ease of exposition because it has a very small share compared to the other store formats

format differentiation in equilibrium and thus mitigate the entry deterrent effect of zoning significantly. In fact, for some ranges of zoning restrictions, the number of firms that enter may be 50 % more than when firms can differentiate on formats relative to when they cannot.

Second, for large ranges of zoning restrictions, which limit the ability to spatially differentiate, there may be no changes in the number of firms in the market, but only changes in the retail mix of formats. This has implications for empirical work, because one may see little changes in entry in response to zoning restrictions and thus may misinterpret the result as zoning having no impact on retailer choices, especially when retailers can differentiate on formats. We have also shown that for any given area available for retail entry, the spatial distribution of the available locations matters for the type of store formats in the market. Specifically, centralized zoning allows for greater format variety, while outskirt zoning leads to lower format variety, with neighborhood zoning at an intermediate level of format variety. This insight on the link between zoning, spatial differentiation and format differentiation is not only important for retailers, but also for city planners who seek to encourage retail format variety in their markets through their zoning authority.

Finally, we find that ignoring spatial zoning regulations in estimating entry and location models, causes serious underestimation of the impact of market characteristics like population and income on store profit potential. It also leads to bias in the true intensity of spatial competition between rivals. The net effect of these biases is that retailers' willingness to enter a market and their propensity to differentiate on formats are underestimated.

We next discuss some key limitations in this paper that warrant future research. First, we abstract away from the fact that entry and location decisions have been made over time and treat entry and location decisions within a static equilibrium framework. A dynamic analysis requires better data (timing of entry and exits). The dynamic analysis becomes practically infeasible in modeling spatial differentiation at a micro-level of 1 square miles, because of the explosion in state space. However given the fineness of zoning regulations and the need to model spatial differentiation carefully, such a detailed modeling of location choices becomes critical. Second, we have treated store entry decisions across markets as independent, unlike recent work by Jia (2008), who models the chain entry decision, taking into account the interdependence across markets. However, her modeling approach is restricted to a small number of competing chains and is hard to extend to our grocery market setting that involves a large number of players. These important issues await future research.

Appendix A

Step 0: ***Initial Population*** - Generate a set of T vectors of starting values for retailers' beliefs about rivals' CCPs for location choices, $\left[\overline{P}_0^1; \overline{P}_0^2; \ldots; \overline{P}_0^T\right]$

Step 1: ***Locally Contractive, q-NPL Iteration*** - Maximize the pseudo likelihood (Eq. 11) to obtain a set of T vectors of parameter estimates: $\Theta_n^t =$

$\arg\max_{\Theta}\left(L\left(\overline{P}_{n-1}^{t},\Theta\right)\right)$, and a new population of CCPs using the q-NPL operator: $\widehat{\overline{P}}_{n}^{t}=\Lambda^{q}\left(\overline{P}_{n-1}^{t},\Theta_{n}^{t}\right)$.

Within each market, normalize the CCPs for each store format so that the CCPs of all formats add up to one. Essentially, for each format *f*, and market location *l*, we have:

$$\widehat{\overline{P}}_{fln}^{t}=\Lambda^{q}\left(\overline{P}_{fln-1}^{t},\Theta_{n}^{t}\right)\Big/\sum_{f=1}^{F}\sum_{l=1}^{l_m}\Lambda^{q}\left(\overline{P}_{fln-1}^{t},\Theta_{n}^{t}\right)$$

Step 2: ***Selection of Parents*** - Based on their fitness, draw, with replacement, T '*mother*' CCP vectors and T '*father*' CCP vectors from the set, $\left[\widehat{\overline{P}}_{n}^{1};\widehat{\overline{P}}_{n}^{2};\ldots;\widehat{\overline{P}}_{n}^{T}\right]$ and form couples or *Parents*. CCPs with high likelihood values, $L\left(\widehat{\overline{P}}_{n}^{t},\Theta_{n}^{t}\right)$, and those closer to convergence (Absolute value of $\left(\widehat{\overline{P}}_{n}^{t}-\widehat{\overline{P}}_{n-1}^{t}\right)$ closer to zero) are considered more fit to continue. In our problem, we use the following *fitness criterion*:

$$h\left(\widehat{\overline{P}}_{n}^{t}\right)=\lambda_1\ln\left[L\left(\widehat{\overline{P}}_{n}^{t},\Theta_{n}^{t}\right)\right]-\lambda_2\left\|\widehat{\overline{P}}_{n}^{t}-\overline{P}_{n-1}^{t}\right\|$$

where, λ_1 and λ_2 are small positive constants. The t^{th} CCP vector gets selected with the probability:

$$S^{t}=\exp\left(h\left(\widehat{\overline{P}}_{n}^{t}\right)\right)\Big/\sum_{j=1}^{T}\exp\left(h\left(\widehat{\overline{P}}_{n}^{j}\right)\right)$$

Now, we have the set of couples: $\left[\left(\widehat{\overline{P}}_{n}^{1'},\widehat{\overline{P}}_{n}^{1''}\right);\left(\widehat{\overline{P}}_{n}^{2'},\widehat{\overline{P}}_{n}^{2''}\right);\ldots;\left(\widehat{\overline{P}}_{n}^{T'},\widehat{\overline{P}}_{n}^{T''}\right)\right]$

Step 3: ***Crossover and Mutation*** - Obtain an *offspring* from each couple as follows:

$$\overline{P}_{n}^{t}=D\cdot\left(\widehat{\overline{P}}_{n}^{t'}+Z_n\cdot\delta_n\cdot\widehat{\overline{P}}_{n}^{t'}\right)+(1-D)\cdot\left(\widehat{\overline{P}}_{n}^{t''}+Z_n\cdot\delta_n\cdot\widehat{\overline{P}}_{n}^{t''}\right)$$

where, D is a vector of indicators for the identity of the parent who provides each element of the CCPs. Its elements are i.i.d. with $\Pr\left(D_j=1\right)=0.5$ for the j^{th} element. Z_n is another vector of indicators for the identity of the elements of the CCPs which undergo mutation. Its elements are also i.i.d. with $\Pr\left(Z_{jn}=1\right)=0.5/\sqrt{n}$. Hence, with multiple iterations, as we get closer to the global optimum, we allow the amount of mutations to reduce to zero. Finally, δ_n is a vector whose elements represent the magnitude of a mutation. It is also defined such that its elements go to zero with multiple iterations. Specifically, we use: $\delta_{jn}\in U(-0.5/\sqrt{n},\tilde{\ }0.5/\sqrt{n})$

As with Step 1, within each market, again normalize the CCPs so that the CCPs of all formats add up to one. Now, we have the new set of CCPs, $\left[\overline{P}_{n}^{1};\overline{P}_{n}^{2};\ldots;\overline{P}_{n}^{T}\right]$.

Iterate Steps 1–3 until the set of CCPs converges.

Appendix B

In their classification of land types, *NLCD 2001* combines high density residential land and commercial land but *NLCD 1992* separates them. Hence, we match the two data sets using *ArcGIS* software to separate the pixel data for all residential land areas from land areas with commercial activity in 2001. We are able to do this separation because land areas which were high density residential in 1992 are unlikely to convert to commercial land areas by 2001, and vice versa. If there is a situation where an area that was low-density residential in 1992 has been classified as commercial land in the 2001 data then we do a quick visual inspection of the geographical area using *Google Earth* to confirm whether that area is truly commercial land or if it has converted into a high density residential land.

References

Aguirregabiria, V., & Mira, P. (2005). "A genetic algorithm for the structural estimation of games with multiple equilibria." Working Paper.

Aguirregabiria, V., & Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica, 75* (1), 1–53.

Bajari, P., Benkard, L., & Levin, J. (2007). Estimating dynamic models of imperfect competition. *Econometrica, 75*(5), 1331–1370.

Bar-Isaac, H., Caruana, G., & Cunat, V., (2009). "Costly search and design." Working paper.

Bresnahan, T., & Reiss, P. (1991). Entry and competition in concentrated markets. *Journal of Political Economy, 99*, 977–1009.

Chung, L. L. W. (1994). The economics of land-use zoning: a literature review and analysis of the work of Coase. *The Town Planning Review, 65*(1), 77–98.

Ciliberto, F., & Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica, 77*(6), 1791–1828.

Datta, S., & Sudhir, K. (2012). "The agglomeration-differentiation tradeoff in spatial location choice." Working paper.

Draganska, M., Mazzeo, M., & Seim, K. (2009). Beyond plain vanilla: modeling joint product assortment and pricing decisions. *Quantitative Marketing and Economics, 7*(2), 105–146.

Ellickson, P. B., Houghton, S., & Timmins C. (2010). "Estimating network economies in retail chains: a revealed preference approach." NBER Working Paper 15832.

Fischel, W. K. (2004). An economic history of zoning and a care for its exclusionary effects. *Urban Studies, 41*(2), 317–340.

Fox, J. (2007). Semiparametric estimation of multinomial discrete choice models using a subset of choices. *The RAND Journal of Economics, 38*(4), 1002–1019.

Fox, J., & Bajari, P. (2010). "Measuring the efficiency of an FCC spectrum auction." Working Paper.

Hanushek, E. A., & Jackson, J. E. (1977). *Statistical methods for social scientists*. New York: Academic.

Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. (2004). Development of a 2001 national land-cover database for the United States. *Photogrammetric Engineering and Remote Sensing, 70*(7), 829–840.

Jia, P. (2008). What happens when Wal-Mart comes to town: an empirical analysis of the discount retailing industry. *Econometrica, 76*(6), 1263–1316.

Kasahara, H., & Shimotsu, K. (2008). "Sequential estimation of structural models with a fixed point constraint." Working Paper.

Kuksov, D. (2004). Buyer search costs and endogenous product design. *Marketing Science, 23*(4), 490–499.

Ladd, B. (1990). "Urban planning and civic order in Germany, 1860–1914." *Harvard Historical Studies*, Harvard University Press.

Mazzeo, M. (2002). Product choice and oligopoly market structure. *The RAND Journal of Economics, 33*, 221–242.

Munro, W. B. (1931). A danger spot in the zoning movement. *The Annals of the American Academy of Political and Social Sciences, 155*(May), 202–206.

Nishida, M. (2010). “Estimating a model of strategic store network choice: The convenience-store industry in Okinawa,” *NET Institute Working Paper* No. 08-27.
Orhun, Y. (2012). “Spatial differentiation in the supermarket industry: The role of common information,” *Quantitative Marketing and Economics,* in this issue.
Pakes, A., Ostrovsky, M., & Berry, S. (2007). Simple estimators for the parameters of discrete dynamic games, with entry/exit examples. *The RAND Journal of Economics, 38*, 373–399.
Pesendorfer, M., & Schmidt-Dengler, P. (2008). Asymptotic least squares estimators for dynamic games. *Review of Economic Studies, 75*, 901–928.
Pogodzinski, J. M. (1991). The effects of fiscal and exclusionary zoning on household location: a critical review. *Journal of Housing Research, 2*(2), 145–160.
Ridley, D. B., Sloan, F. A., & Song, Y. (2010). “Retail zoning and competition.” Working Paper.
Seim, K. (2006). An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics, 37*(3), 619–640.
Su, C., & Judd, K. L. (2012). Constrained optimization approaches to estimation of structural models. *Econometrica, 80*(5), 2213–2230
Suzuki, J. (2010), “Land use regulation as a barrier to entry: Evidence from the Texas lodging industry.” Working Paper.
Thomadsen, R. (2005). The effect of ownership structure on prices in geographically differentiated industries. *The RAND Journal of Economics, 36*(4), 908–929.
Vitorino, M. A. (2010). “Empirical entry games with complementarities: An application to the shopping center industry.” Working Paper.
Vogelmann, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K., & Van Driel, J. N. (2001). Completion of the 1990’s national land cover data set for the conterminous United States. *Photogrammetric Engineering and Remote Sensing, 67*(6), 650–662.
Watson, R. (2009). Product variety and competition in the retail market for eyeglasses. *The Journal of Industrial Economics, 57*(2), 217–251.
Zhu, T., & Singh, V. (2009). Spatial competition with endogenous location choices: an application to discount retailing. *Quantitative Marketing and Economics, 7*(1), 1–35.
Zhu, T., Singh, V., & Manuszak, M. (2009). Market structure and competition in the retail discount industry. *Journal of Marketing Research, 46*(4), 453–466.