

Estimating Network Economies in Retail Chains: A Revealed Preference Approach

Paul B. Ellickson* Stephanie Houghton
University of Rochester Texas A&M University

Christopher Timmins
Duke University and NBER

January 29, 2013

Abstract

We measure the effects of chain economies, business stealing, and heterogeneous firms' comparative advantages in the discount retail industry. Traditional entry models are ill-suited for this high-dimensional problem of strategic interaction. Building upon recently developed profit inequality techniques, our model admits any number of potential rivals and stores per location, an endogenous distribution network, and unobserved (to the econometrician) location attributes that may cause firms to cluster their stores. In an application, we find that Wal-Mart benefits most from local chain economies, while Target shows a greater ability to respond to rival competition. Kmart exhibits neither of these strengths. We explore these results with counterfactual simulations highlighting these offsetting effects and find that local chain economies play an important role in securing Wal-Mart's industry leader status.

*Ellickson: Simon Graduate School of Business Administration, University of Rochester, Box 270100 Rochester, New York 14627 (email: paul.ellickson@Simon.Rochester.edu); Houghton: Department of Economics, Texas A&M University, 4228 TAMU, College Station, TX 77843 (email: houghton@econmail.tamu.edu); Timmins: Department of Economics, Duke University, PO Box 90097, Durham, NC 27708 (email: christopher.timmins@duke.edu). The authors would like to thank Arie Beresteanu, Tim Bresnahan, Jeremy Fox, Paul Grieco, Tom Holmes, Shakeeb Khan, Sanjog Misra, and Andrew Sweeting for their helpful comments. We have also greatly benefited from the thoughtful suggestions of participants in the NBER Winter IO Meetings, Duke University Applied Microeconomics Lunch, the Minnesota Applied Micro Workshop, the Triangle Econometrics Conference, and seminar participants at Rochester, Texas A&M, the University of Texas-Austin, Wisconsin, the University of Texas at Dallas, Virginia Tech, and the Yale SOM. Any remaining errors and omissions are our own.

1 Introduction

Over the past few decades, “big box” chain stores have come to dominate many sectors of retail trade. While the chain store format dates back to the early 1900s, the “big box” phenomenon is more recent, relying on the diffusion of the automobile and the great American migration to the suburbs. Just as the much larger supermarket format replaced corner grocers in the 1950s, “big box” superstores replaced smaller “mom n’ pop” sporting good, book, and hardware stores in the 1980s and 1990s. The leading example is Wal-Mart, which alone accounted for 9.6% of U.S. retail sales in 2008, excluding automobile and gasoline sales. In the discount retail sector specifically, Wal-Mart, Kmart and Target now comprise three-quarters of total sector sales and show few signs of slowing down. The chain store advantage is both national and local: chains take advantage of scale economies in purchasing and logistics, as well as more local “density” economies that arise from sharing local knowledge and pooling advertising resources. Despite the increasing dominance of chain superstores, there is relatively little empirical work quantifying the size and sources of the chain advantage.¹ This is the goal of our paper.

In part, the lack of empirical research reflects the scale and complexity of the chain store’s decision problem: building a massive network of outlets and distribution centers that move products to consumers as efficiently as possible, recognizing that their rivals are doing the same. Until recently, a complex strategic game on such a high dimensional space was thought impossible to solve, much less estimate.² Furthermore, many retail sectors are dominated by as few as two or three firms, begging the question of how one can recover rich systems of parameters with the observed behavior of so few players. Building on the profit inequalities approach developed by Pakes et al. (2011) and Bajari et al. (2007), we propose a scalable empirical framework for analyzing network entry problems that does not require solving for an equilibrium, or even that the equilibrium be unique or played in pure strategies. Rather, it simply requires enforcing the equilibrium condition that the observed

¹Jia (2008), Holmes (2011) and Nishida (2008) are three notable exceptions, discussed in detail below. Earlier, Smith (2004) applied discrete choice demand estimation to analyze mergers between supermarket chains.

²Jia (2008) exploited the lattice structure of a two-firm game to reduce the computational burden of solving for equilibria, while Fox (2007) and Bajari and Fox (2011) use the tools of maximum score and revealed preference to avoid solving for choice probabilities at all.

choices yield weakly higher expected payoffs than any feasible alternatives, holding rivals’ actions fixed. We apply our estimator to the discount retail sector, isolating the comparative advantages of each of the three players and quantifying the relative importance of the local chain effect. Moreover, given the relative simplicity of our estimation approach, this framework could easily be applied in other multi-store or multi-product settings such as airline hub and spoke networks, automobile product line decisions, or retailer assortment problems.

Our empirical framework builds on the profit inequalities (henceforth, PI) framework developed in Pakes et al. (2011) and Bajari et al. (2007), incorporating the pairwise maximum score approach to large scale choice problems proposed by Fox (2007) and implemented by Bajari and Fox (2013). Rather than attempting to fully solve a complex discrete game for empirical choice probabilities that are then matched to data³, the PI approach relies on a revealed preference argument that involves comparing actual choices to a large set of counterfactual actions that each firm could have taken but elected not to. Since the actual decisions arose from equilibrium play, these counterfactual choices must be payoff reducing, creating a large set of profit inequalities on which to base estimation. By relying on profit inequalities rather than equilibrium choice probabilities, the PI approach avoids solving for equilibria, mitigating concerns over multiplicity and eliminating the need to impose (or estimate) a selection rule. This simplified structure makes the framework scalable: it can handle any number of players, subject only to memory and data constraints. In contrast to the alternative choice probability framework, a complex choice space works to the advantage of the PI approach, providing more inequalities with which to pin down parameter estimates.

In practice, we consider a large number of “local” perturbations to the observed chain-network structure that involve swapping pairs of stores owned by rival firms between matched pairs of markets. Along with creating the profit inequalities on which we base

³This is the approach utilized in the “generalized discrete choice” (Pakes (2010)) framework proposed by Tamer (2003) and extended by Ciliberto and Tamer (2009), Andrews and Jia (2012), and Beresteanu et al. (2011). In these settings, *partial identification* often obtains due to the possibility of multiple equilibria: the model can only provide bounds on the range of permissible choice probabilities. Alternatively, the researcher can impose or estimate a selection rule to restore point identification (Bajari, Hong, and Ryan (2010)), although this still requires solving for all the equilibria. Partial identification can also arise in the PI framework if certain support conditions do not hold. These conditions are discussed in detail below.

estimation, this swapping of matched pairs serves a second purpose: eliminating any common sources of unobserved market heterogeneity. In particular, by considering only *matched* swaps of rival stores between pairs of markets, these *difference-in-difference* inequalities allow us to eliminate a common, market-level unobservable, which is key to obtaining unbiased estimates of local chain and competition effects. Because we control for the structural error in this way, and because the markets from which stores are swapped are chosen randomly, the profit inequalities satisfy a rank-ordering property and can be estimated using pairwise maximum score (Fox (2007)), a conditional *median* inequality analog of the moment inequality methods proposed by Pakes et al. (2011).

Intuitively, the PI approach exploits the complexity of the decision space to offset the small number of market participants noted above. Even though the number of players is small (in this case, three), the number of possible alternative configurations (and pairwise deviations) is immense, allowing us to exploit asymptotics in the number of choices, rather than the number of agents. However, the small number of players does impose a cost when it comes to recovering and decomposing the fixed effects that dropped out of the difference-in-difference inequalities. Doing so is critical for using the parameter estimates to learn about marginal effects of specific common drivers of firm behavior (e.g., population) or for simulating counterfactual equilibria resulting from a change in the underlying primitives. Employing another set of inequalities, based this time on the first order conditions from the profit maximization problem, we recover ranges in which these fixed effects can lie, and then project these intervals onto additional covariates using set inference techniques developed by Beresteanu and Molinari (2008). Set identification follows from the limited number of inequalities available given the small number of players observed in this industry.

We apply our estimator to the store location choice problem faced by Wal-Mart, Kmart, and Target. Treating the decision of where (and how many stores) to locate in each of 1351 markets as a static discrete game, we find that each firm has individual preferences and unique strengths. Consistent with conventional wisdom, we find that Wal-Mart favors rural locations, Kmart smaller urban locales, and Target more urban settings. We find that all firms impose strong competitive pressures on one another, but that each firm (particularly Wal-Mart) does exploit some degree of local density economies (i.e., a “local chain” effect). Notably, to offset the presence of an additional store, Wal-Mart requires the smallest increase

in market population if that new store is another Wal-Mart. Target requires the smallest increase if that store belongs to one of its rivals. Kmart, on the other hand, is both unable to capitalize on local chain economies or fend off its rivals — a result that explains its poor market performance.

In a counterfactual simulation of a new equilibrium in which local chain effects are “turned off” we find a drastic reduction in entry by Wal-Mart. Target takes advantage of this reduced competition by dramatically increasing its own entry behavior (particularly in large markets). The absence of agglomeration effects and the increased competition from Target lead to a modest reduction in entry by Kmart. The implication of this counterfactual is that an important part of Wal-Mart’s market dominance arises from its ability to exploit local chain economies.

Our paper builds on and extends the static entry literature that started with Bresnahan and Reiss (1990, 1991) and Berry (1992).⁴ We also contribute to a growing literature on Wal-Mart itself, which includes Basker (2005, 2007), Ellickson and Grieco (2013), Ellickson, Misra, and Nair (2012), Hausman and Leibtag (2007), Neumark, Zhang, and Ciccarella (2008), Zhu and Singh (2009), and Zhu, Singh, and Manuszak (2009).

While most authors have focused on isolated entry decisions by single-product firms, the first structural papers to address the chain store network problem directly are Jia (2008), Nishida (2008), and Holmes (2011). Jia (2008) examines the strategic store placement decisions of Kmart and Wal-Mart, using a full solution method that exploits a supermodular representation of the two-player game. Her elegant, lattice-based solution mechanism comes at a cost, restricting her analysis to only two national players and requiring that the net chain effects be non-negative (i.e. that business stealing not outweigh the benefits of agglomeration). She models network economies as positive spillovers occurring across nearby locations (defined as counties), inferring the chain effect from the extent to which firms cluster their own stores. However, she is only able to consider counties that contain a single store per firm. Nishida (2008) extends Jia’s approach to allow any number of stores to be placed in a location, but still requires the model to exhibit positive spatial spillovers across locations and can only accommodate two competing firms. Both approaches require

⁴Other notable papers include Mazzeo (2002), Tamer (2003), Seim (2006), Akerberg and Gowrisankaran (2006) and Ciliberto and Tamer (2009). Ellickson and Misra (2011) provide an overview of this growing literature.

imposing an equilibrium selection rule prior to estimation. Our analysis relaxes these restrictions, so that we may consider the full set of potential locations, allow for any number of competing firms, and place no restrictions on the number of stores per location or the net impact of spatial spillovers. We are also able to control for proximity to endogenously placed distribution centers and remain agnostic with regard to equilibrium selection.

Holmes (2011) examines the spatial structure of Wal-Mart’s national network from the perspective of dynamic discrete choice, using Wal-Mart’s sequential decisions over where to open additional stores to infer the importance of economies of density. In particular, the degree to which Wal-Mart passed up early entry into desirable locations to build additional stores closer to a distribution center reveals their preference for such proximity. Holmes also incorporates information on store-level revenues, further identifying density economies from the extent to which Wal-Mart tolerates cannibalization of sales by its own nearby stores. However, in acknowledging the dynamic aspects of Wal-Mart’s decision process, Holmes does not model (or control for) reactions to (or by) Wal-Mart’s rivals. Like Jia, we abstract from the dynamics of the chain problem to focus on strategic interaction.⁵ However, in contrast to both Jia and Holmes, our identification strategy turns on the degree to which firms prefer to face another one of their own outlets rather than an additional outlet of a rival.

The remainder of this paper proceeds as follows. Section 2 presents details about the data and provides background on the discount store industry, which motivates our modeling and estimation approach. Section 3 presents a model of firm profit as a function of the attributes of markets where they place their stores. We then outline a three-part estimation algorithm used to recover these payoff functions. In the first stage, we estimate the strategic components of store-level profits via pairwise maximum score. In the second and third stages, we recover set-valued estimates of the common, market-level fixed effect and decompose that fixed effect to estimate the remaining components of firm profit. Section 4 presents the empirical results and further explores their implications by applying them to a

⁵While the dynamic aspects of retail competition are clearly important, incorporating them into a strategic model of competition is greatly complicated by the non-stationary structure of the discount store industry. Wal-Mart has never exited a market and Kmart hasn’t opened a new store in over six years. While there are some signs that the industry has stabilized in recent years, the available data currently reflect the process of expanding to equilibrium, as opposed to the stationary ergodic structure typically assumed for dynamic games. See Beresteanu and Ellickson (2006) for an attempt to address some of these issues.

counterfactual scenario in which the positive benefits of clustering chain stores are removed. Section 5 concludes.

2 The Discount Store Industry

While the segment of retail now served by chain discount stores was once quite fragmented, in the past few decades it has come to be dominated by three main players: Wal-Mart, Kmart, and Target. While each firm is essentially national in scope, there are some fairly obvious distinctions in the types of markets they choose to serve, partly driven by the types of consumers they target. Consistent with its rural beginnings and choice of merchandise, Wal-Mart favors rural locations and smaller cities, avoiding the largest cities almost completely (not always by choice). Target, on the other hand, prefers urban locations, consistent with its trendier, “up-market” focus. Finally, Kmart is somewhat less focused, having stores both in cities and rural towns, but with an overall preference for smaller urban locations. In addition to proximity to consumers, location choice is also constrained by logistical concerns - stores have to be stocked with merchandise from a regional distribution center, and so may benefit from being close to that hub and close to one another. These economies of density may also come in the form of shared advertising, shared local market knowledge, and managerial spillovers between stores in the same market. All three firms balance these economies of scale and density against the idiosyncratic preferences of individual consumers by designing an optimal network of stores.

Before verifying these stylized facts and motivating our specific modeling choices, we first introduce our dataset and choose a sensible definition of a “local market”. In retail industries, location choice essentially dictates the types of consumers you are going to serve, as individuals tend to sort themselves into reasonably homogeneous neighborhoods. While such sorting could (and probably does) occur on a very local level, from the researcher’s perspective, there is a trade-off between choosing a fine geographic grid and obtaining meaningful demographic covariates. With this trade-off in mind, we chose Core Based Statistical Areas (CBSAs) as our basic building block, which we will call a market or location. CBSA refers collectively to metropolitan statistical areas and smaller “micropolitan” statistical areas. Firms choose which locations to enter and how many stores to build in each. While

this coarse-grained market definition ignores more nuanced aspects of spatial differentiation, it allows us to account for correlated, market-level unobservables, which are key to correctly identifying network and competition effects. In addition to these local characteristics, an important determinant of store location choice is the distance to primary suppliers, which we capture using detailed information on distribution centers unique to our dataset.

The data on discount stores are drawn from the *Trade Dimensions* Retail Tenant Database for 2006. Both the dataset and the construction of local markets are described in detail in Appendix A. Figure 1 illustrates the locations of all the discount stores in our data set, along with the locations of all distribution centers. Target’s tendency to cluster in urban locations is clear, providing a sharp contrast with the more uniform distribution of Wal-Mart and Kmart locations. Figure 2 focuses in on the upper Midwest, bringing this contrast into sharper focus. Note that Target tends to place distribution centers close to large cities, which is consistent with supporting its generally accepted upscale, urban focus. Wal-Mart, on the other hand, places its distribution centers so as to best serve stores which are themselves located in the interstices between major markets.

Overall, Wal-Mart has a presence in 97% of these markets, compared to 44% and 26% of the markets for Kmart and Target, respectively (see Tables 5 and 6 in Appendix A). In part, this is expected because there are more than twice as many Wal-Mart stores included in the data (3345) as there are Kmart or Target stores (1362 and 1443, respectively). However, it also appears that these firms follow quite different strategies when locating stores. This is illustrated clearly in Table 6, which summarizes the location characteristics for each store. Over 95% of Target’s stores are located in metropolitan statistical areas, compared to only 68% of Wal-Mart’s stores and 78% of Kmart’s stores.

Finally, Figure 3 examines the relationship between the number of discount stores located in each market and population. The top left panel, which is in levels and includes the fitted regression line from a simple quantile (median) regression, reveals that the relationship is fairly linear, at least up until the very largest markets (Los Angeles and New York City), indicating that population is the primary driver of demand. The top right panel includes only those markets with population under 500,000, along with the fitted values from the median regression. Again, the relationship is linear, and fairly tightly clustered around the regression line at these low values of population. Following Bresnahan and Reiss (1990),

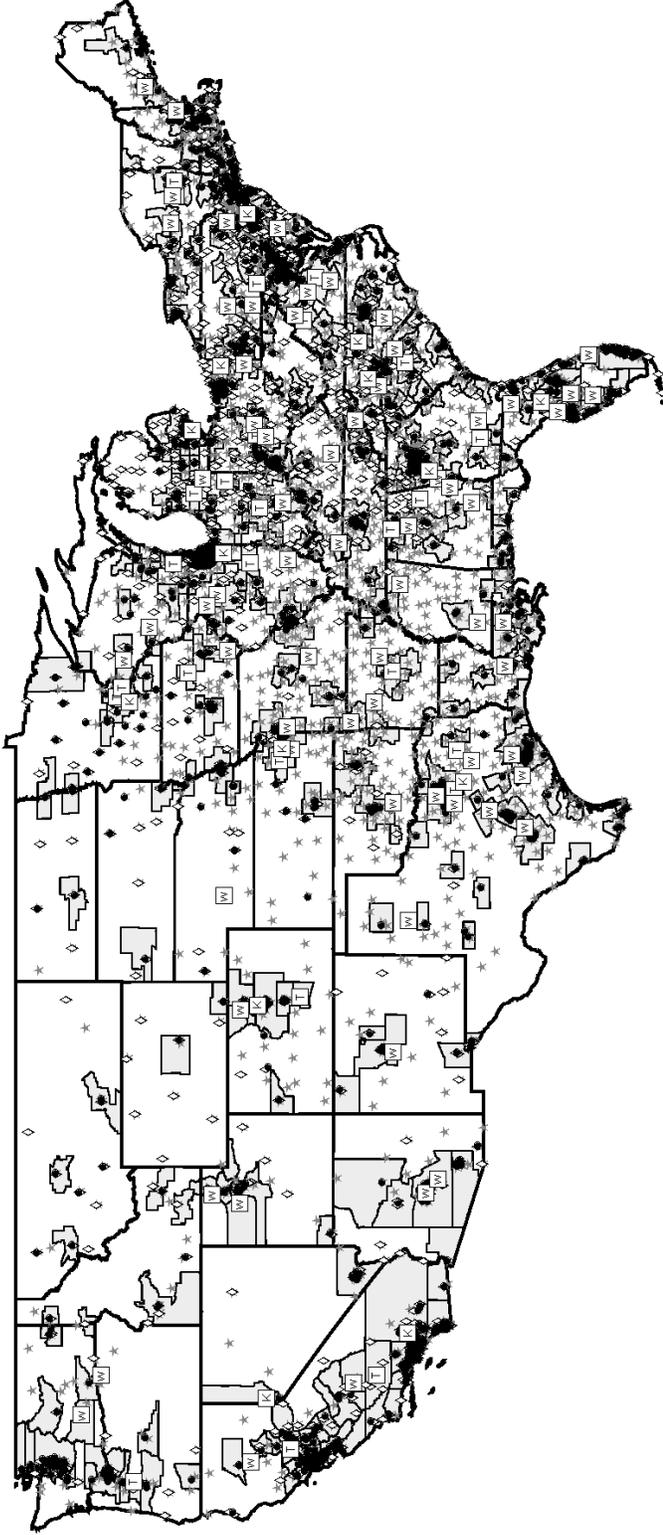
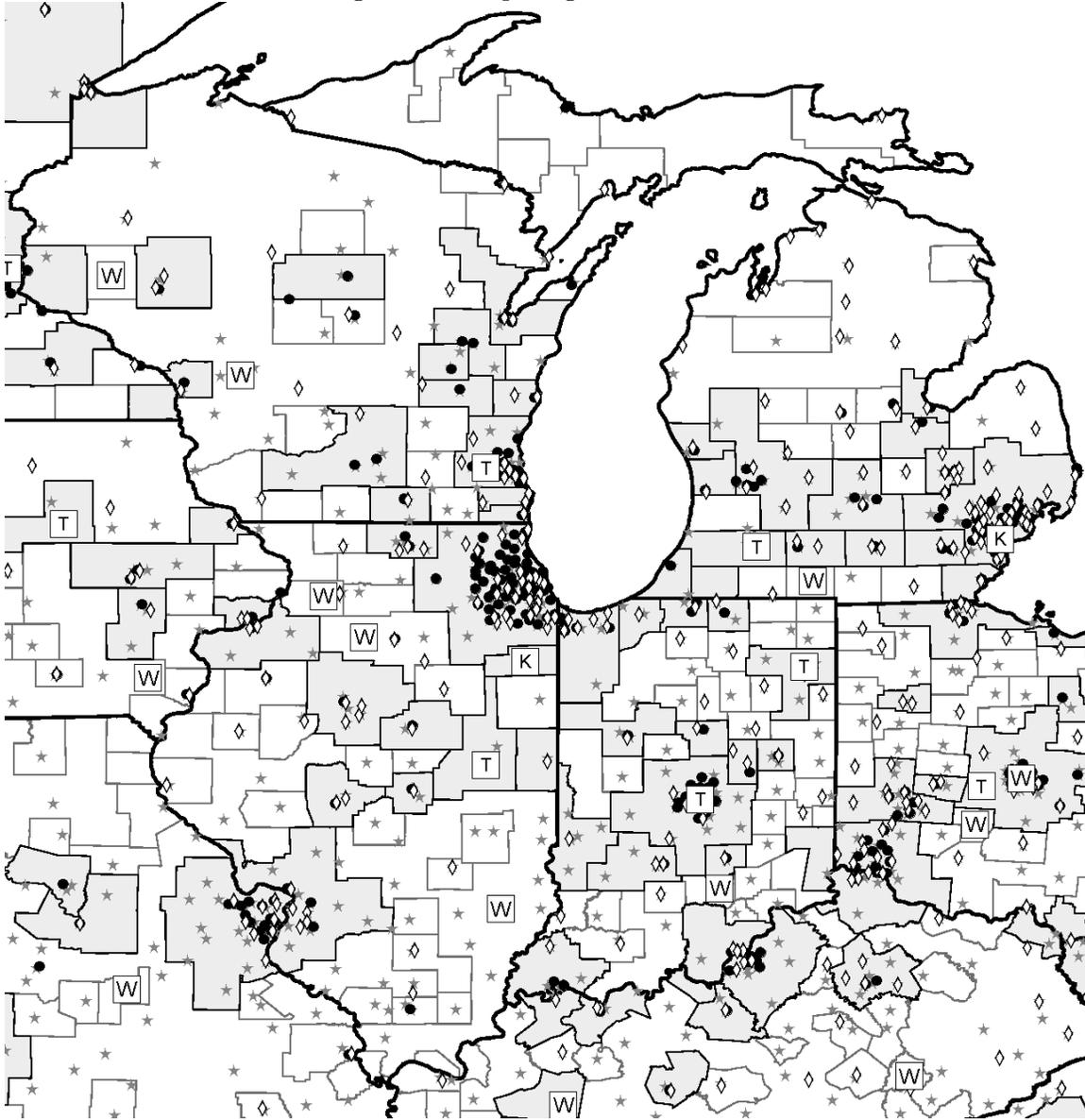


Figure 1: The Discount Store Industry

Target stores are denoted by solid black circles, Wal-Mart stores are denoted by gray stars, and Kmart stores are denoted by white diamond shapes. Distribution centers appear as squares inset with the operating firm's initial. The shaded regions indicate the boundaries of metropolitan statistical areas. Micropolitan statistical areas and isolated counties, which also serve as markets in our model, have been left off this particular map to improve clarity.

Figure 2: Competing in the Midwest



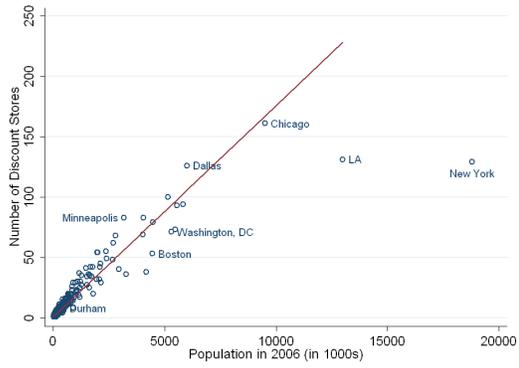
Target stores are denoted by solid black circles, Wal-Mart stores are denoted by gray stars, and Kmart stores are denoted by white diamond shapes. Distribution centers appear as squares inset with the operating firm's initial. The shaded regions indicate the boundaries of metropolitan statistical areas, while the unshaded regions indicate smaller micropolitan statistical areas.

we use order statistics from markets that contain only a single store to provide a ballpark estimate of the minimum efficient scale (MES), in population units, that is required to support the first entrant. The median population (26,450) and interquartile range (18,420; 38,240) for markets that contain just one store suggest that it takes about 25,000 people to induce the initial entry.

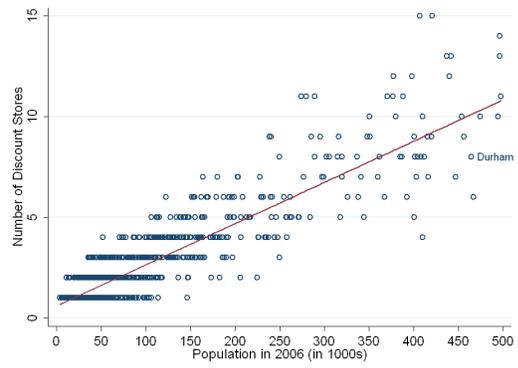
Nonetheless, there are several markets whose population is well below this critical level and yet contain one or more stores, and several more whose population is well above it yet contain just a single store (note that we are only plotting those markets that contain at least one store). Undoubtedly, the small markets with multiple outlets contain unobserved features (e.g. proximity to an interstate or highway interchange) that draw consumers from a larger geographic basin, while the larger markets with relatively few stores have offsetting unobservables (e.g. higher land prices and tighter zoning restrictions) that dampen entry. This heterogeneity, especially at the lowest levels of population, illustrates clearly why we need to control for market-level unobservables. There is often little in the data to explain why a market with 40,000 people has three discount stores but a market with 80,000 has just one. If we simply conditioned on the presence of own or rival stores, we would be biased toward finding a preference for agglomeration. This is exactly what we find below.

The bottom left panel returns to the full set of markets but switches the relationship from levels to logs, revealing a fairly tight, but heteroskedastic linear relationship (in logs). The fitted value line, which is from a restricted cubic spline regression, demonstrates a sharp kink near the level of $\ln(\text{population})$ where we would expect to see a second store ($\ln(2 \times 25000) \approx 10.8$), reflecting the inherent indivisibility of the big box format. A similar relationship holds firm by firm, as is demonstrated in the lower right panel, which contains fitted values from restricted cubic spline regressions for each firm. This obvious non-linearity further motivates our use of non-parametric fixed effects to control for features of the market which are common to all three firms. These empirical features will also motivate our choice of functional form in the following structural model of entry.

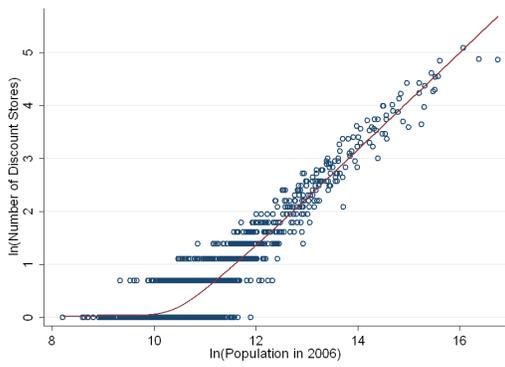
Figure 3: Stores versus Population



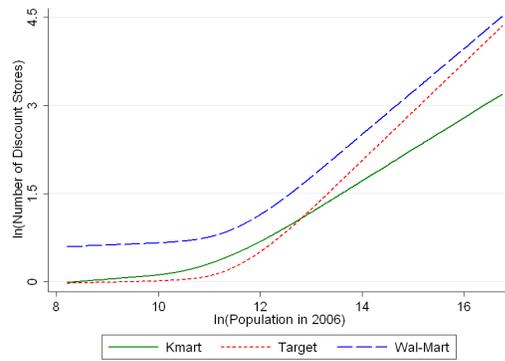
(a)



(b)



(c)



(d)

3 Model

Since we do not have information on prices and quantities (or detailed store characteristics), we follow the standard convention of the entry literature in assuming that each firm’s profit function is known up to a vector of parameters. We have chosen a functional form that is flexible in the number of own stores in order to accommodate *net* “chain” effects that may be either positive or negative. We treat the strategic setting as a game of complete information, in which the econometrician observes covariates that may differ (in a median independent way) from the exogenous variables upon which the firms based their decisions. Consistent with the determinants of behavior described above, we begin by modeling the *per-store* payoff to firm $f = \{T, K, W\}$ of each store in market j as:

$$\begin{aligned} \pi_j^f &= \beta^{f,Own} \ln(N_j^f + 1) + \beta^{f,Own2} (\ln(N_j^f + 1))^2 + \beta^{f,Own3} (\ln(Pop_j)) (\ln(N_j^f + 1)) \\ &\quad + \beta^{f,Other} \ln(N_j^{-f} + 1) + \beta^{f,HQ} \ln(HQ_j^f) + \beta^{f,DC} \ln(DC_j^f) + \beta^{f,X} X_j + \theta_j \end{aligned}$$

where N_j^f is the total number of stores firm f operates in market j , N_j^{-f} is the total number of rival stores that firm f faces in market j , DC_j^f is the distance from market j to firm f ’s nearest distribution center, HQ_j^f is the distance from market j to firm f ’s headquarters,⁶ X_j is a vector of exogenous attributes of market j (e.g., income, population, household size, regional dummies), β is a vector of parameters with $\beta^{W,X} = 0 \forall X$, and θ_j is a market-specific fixed effect that is common to all firms. We further decompose this fixed effect as a combination of observable and unobservable market characteristics:

$$\theta_j = f(X_j; \gamma^X) + \xi_j$$

where ξ_j is an additively separable unobserved (to the econometrician) attribute of market j , that is assumed to be scalar, common across firms, and uncorrelated with X_j . Note that $f(X_j; \gamma^X)$ captures the way in which the X_j variables affect Wal-Mart’s profits and is also important for determining how these variables impact the profits of the other two firms. In particular, if $f(X_j; \gamma^X)$ were a simple linear function $\gamma^X X_j$, the impact of X_j on Kmart’s

⁶As Holmes (2011) demonstrates, Wal-Mart’s outward expansion began from its headquarters in Bentonville, Arkansas. Target has had a historical presence near its parent-company (Dayton-Hudson) headquarters in Minneapolis, Minnesota. Kmart’s ties to its original headquarters in Troy, Michigan are less apparent in its spatial layout.

profits could be represented by $(\beta^{K,X} + \gamma^X)X_j$, while the effect of X_j on Target's profits would follow from a comparable expression.

We expect to find $\beta^{f,DC}, \beta^{f,HQ} < 0$, the former providing evidence that costly transportation creates incentives for firms to place stores closer to their distribution centers, and the latter reflecting other coordination costs and the historical pattern of expansion from a firm's home base. The most important strategic parameters are $\beta^{f,Own}, \beta^{f,Own2}, \beta^{f,Own3}$, and $\beta^{f,Other}$, which capture the net impact of own and rival stores on firm profits. While the impact of rival stores should clearly be negative ($\beta^{f,Other} < 0$), the choice of functional form is designed to allow for flexibility in the effect of nearby own-stores on profits, leaving open the possibility of recovering either net agglomeration or business stealing effects across the range of store counts.

Since ex ante, per-store profits are the same at every store in a given market, calculating total firm-level profits involves summing over all J markets:⁷

$$\Pi^f = \sum_{j=1}^J N_j^f \times \pi_j^f,$$

This is, ultimately, the quantity that firms are maximizing through the choice of N_j and the location of distribution centers (which in turn determines DC_j^f in per-store profits). The derivative of π_j^f with respect to N_j^f reveals the presence of *net* agglomeration or own-business stealing effects. In particular, over the range of N_j^f where $\frac{d\pi_j^f}{dN_j^f}$ is positive, firm f experiences net agglomeration effects in location j (i.e., adding another store raises the marginal profitability of all stores in the location). Over the range where $\frac{d\pi_j^f}{dN_j^f}$ is negative, own-business stealing effects dominate agglomeration effects, and additional stores decrease the profitability of the firm's other stores in the market.

An important feature of our modeling framework is the inclusion of the unobserved market attribute, ξ_j , which serves as our structural error. From a practical standpoint, it is unlikely that our vector of observable market attributes, X_j , will capture everything about a market that is relevant to profitability. If we ignore these unobservable characteristics,

⁷Note that our model only requires a firm's expectation of the profit at each store in a market to be the same at the time of siting. Profits will likely vary across stores after entry. We need only require that firms cannot predict these deviations. While it may be tempting to assume that larger markets should have more profitable stores, the expansion in demand is mostly reflected in the number of outlets, rather than revenue per store (Ellickson, 2007).

we will likely arrive at biased estimates of a number of parameters – particularly those associated with N_j^f . Of course, explicitly including ξ_j as part of the error structure creates potential endogeneity problems since N_j^f is an equilibrium construct: N_j^f is a direct function of ξ_j as well as an indirect function of $\xi_k \forall k \neq j$. In solving this problem, we rely on three identification assumptions. First, the endogenous attributes of each market are firm-specific (i.e., the number of own-stores, stores operated by other firms, and distance to distribution center are different for each firm in each location). Second, the unobserved attribute ξ_j is common across firms and additively separable in the profit function. Third, after controlling for the common unobservable ξ_j , any remaining errors are median independent of N_j^f , N_j^{-f} , and DC_j^f when evaluated at the true parameter vector. The first two assumptions are common in both the sorting and discrete choice (demand) literature. The final assumption is closely related to the error decomposition described in Pakes et al. (2011) and requires further explanation.

To control for the endogeneity of ξ_j , we employ a double differencing procedure that eliminates this unobserved market attribute from the set of inequalities upon which we base our estimation. However, by differencing away the ξ 's, we have eliminated the main stochastic component of the profit functions. Therefore, without an additional source of randomness, the model could not rationalize all potential outcomes (it would be stochastically singular). To address this concern, we assume that the observed outcomes also depend on an additional set of shocks that are median independent of the included covariates.⁸ We view these disturbances as capturing differences between the agents' ex ante beliefs regarding the values of the conditioning variables and the ex post values employed by the researcher. For example, we measure observable demographics like population and income using their 2006 values, but most of the entry decisions we model were taken over the previous 25 years. It seems reasonable to assume that firms may have incorrectly predicted how these variables would evolve over time, making their decisions before the true (recorded) values were ultimately revealed.⁹ Using arguments from Hyslop and Imbens (2001), Mahajan et

⁸In contrast to the alternative *generalized discrete choice* approach (Tamer (2003), Ciliberto and Tamer (2009)), it is difficult to give this error a structural interpretation as an additive component of the profit function; it is more naturally defined in differences (see below). However, it does eliminate the need to specify the content of agents information sets or solve for equilibria conditional on all possible realizations of rival's choices (see Pakes (2010)), the latter of which is key for scalability.

⁹A similar timing assumption is used in Eizenberg (2012).

al. (2011) establish robustness of the maximum score estimator to this form of “prediction error” in the context of a simpler single agent binary discrete choice problem. We conjecture that under possibly stronger conditions, similar results might obtain in our more complex setting as well.¹⁰

Given our modeling strategy, it is important that we include data describing key attributes of each market that might be valued differently by different firms. One obvious candidate is simply geography – the locations of corporate headquarters can explain much of Wal-Mart’s prominence in the South. The only constraint in incorporating additional market-level information in our empirical specification is data availability.

We model the equilibrium locations of all Targets, Wal-Mart’s, and Kmart’s as of 2006.¹¹ The locations of each firm’s distribution centers are important determinants of these siting decisions, but are certainly not exogenously determined. There is a potentially complicated model of firms’ decisions about where to place these distribution centers that we do not attempt to specify. Instead, we deal econometrically with the fact that DC_j^f is endogenous, allowing us to recover unbiased estimates of firm preferences.

3.1 Estimation of Strategic, Firm-Specific Parameters

Our estimator is based on a revealed preference approach that uses pairwise comparisons between the observed location decisions made by firms and specific “single-store” deviations. The assumption is that a single-store deviation (i.e., taking a single store and moving it to a new location, holding the location decisions of other firms fixed) is a deviation from the observed Nash equilibrium strategy and is, therefore, payoff reducing for the firm.

Recall that the per-store profit of a given firm (say Target) in market j is given by

¹⁰It should be noted that this forecast error is distinct from classical measurement error, to which maximum score estimation is not robust, as it averages over a non-linear (indicator) function of the data. A potential advantage of the moment inequality approach over ours is its robustness to classical measurement and specification error, which would allow a broader interpretation of these profit shocks (and a formal proof of consistency). As an additional robustness check, we re-estimated our model using an alternative moment-inequalities-based approach instead, yielding similar results. See footnote 14 for additional discussion.

¹¹Following the standard practice in the static entry literature, we treat the entire network of each firm as being determined simultaneously in a one-shot game. In contrast, Holmes (2011) models the dynamics of store diffusions, albeit from the perspective of only a single firm.

$$\begin{aligned}\pi_j^T &= \beta^{T,Own} \ln(N_j^T + 1) + \beta^{T,Own2} \left(\ln(N_j^T + 1)\right)^2 + \beta^{T,Own3} (\ln(Pop_j)) \left(\ln(N_j^T + 1)\right) \\ &\quad + \beta^{T,Other} \ln(N_j^K + N_j^W + 1) + \beta^{T,HQ} \ln(HQ_j^T) + \beta^{T,DC} \ln(DC_j^T) + \beta^{T,X} X_j + \theta_j\end{aligned}$$

Since the ex ante per-store profit is the same for every store in a given market, Target's total profit in market j can be written

$$\Pi_j^T = N_j^T \cdot (V_j^T(N_j^T) + \theta_j)$$

where

$$\begin{aligned}V_j^T(N_j^T) &= \beta^{T,Own} \ln(N_j^T + 1) + \beta^{T,Own2} \left(\ln(N_j^T + 1)\right)^2 + \beta^{T,Own3} (\ln(Pop_j)) \left(\ln(N_j^T + 1)\right) \\ &\quad + \beta^{T,Other} \ln(N_j^K + N_j^W + 1) + \beta^{T,HQ} \ln(HQ_j^T) + \beta^{T,DC} \ln(DC_j^T) + \beta^{T,X} X_j\end{aligned}$$

Note that N_j^W and N_j^K are implicitly held at their observed values in $V_j^T(N_j^T)$, as we only consider unilateral deviations. The logic of the estimator is to consider “swaps” of a single store from one market to another. For example, consider moving a single Target store from market a (e.g., Minneapolis), which currently contains N_a^T Target stores, to market b (e.g., Chicago), which currently contains N_b^T Target stores. Since the observed configuration (N_a^T, N_b^T) is part of an equilibrium, Target's total profits must be weakly higher under the observed configuration than under the proposed counterfactual configuration $(N_a^T - 1, N_b^T + 1)$. Note that since all “spillovers” (i.e., congestion effects or economies of density) are assumed to occur within, but not across, CBSA markets (i.e., profits are additively separable across markets), the change in total firm profits only depends on the incremental changes associated with the two markets exchanging stores, yielding the following relatively simple profit inequality

$$\begin{aligned}N_a^T \cdot V_a^T(N_a^T) + N_a^T \cdot \theta_a + N_b^T \cdot V_b^T(N_b^T) + N_b^T \cdot \theta_b \\ \geq (N_a^T - 1) \cdot V_a^T(N_a^T - 1) + (N_a^T - 1) \cdot \theta_a + (N_b^T + 1) \cdot V_b^T(N_b^T + 1) + (N_b^T + 1) \cdot \theta_b\end{aligned}$$

Simplifying this expression yields

$$\widetilde{\Delta V}_a^T(N_a^T, N_a^T - 1) + \widetilde{\Delta V}_b^T(N_b^T, N_b^T + 1) + (\theta_a - \theta_b) \geq 0 \quad (1)$$

where the $\widetilde{\Delta V}_j^f(\cdot, \cdot)$ notation represents the decrease (or increase) in profits associated with removing a single store from (or adding a single store to) a particular market.

An impediment to using this simple inequality to recover the structural parameters of Target’s payoff function is the difference in fixed effects, $(\theta_a - \theta_b)$, which, if treated as the econometric error, would clearly be correlated with the included regressors. In addition to the valuation of the exogenous features of the market which are common to all players (represented here as Wal-Mart’s preferences over population, for example), these fixed effects contain the parameters ξ_a and ξ_b , which are common knowledge of the firms, but unobserved by the econometrician. To obtain unbiased estimates of the profit parameters, we must eliminate these common unobservables from the profit inequality. Therefore, we consider another hypothetical store movement that will allow us to difference these fixed effects away. In particular, consider moving one Kmart store from b to a . This yields an inequality similar to the one above, but with an offsetting difference in the fixed effect term:

$$\widetilde{\Delta V}_a^K(N_a^K, N_a^K + 1) + \widetilde{\Delta V}_b^K(N_b^K, N_b^K - 1) + (\theta_b - \theta_a) \geq 0 \quad (2)$$

Adding equation (2) to equation (1) yields another inequality that is free from this problematic term:¹²

$$\begin{aligned} \widetilde{\Delta V}_a^T(N_a^T, N_a^T - 1) + \widetilde{\Delta V}_b^T(N_b^T, N_b^T + 1) \\ + \widetilde{\Delta V}_a^K(N_a^K, N_a^K + 1) + \widetilde{\Delta V}_b^K(N_b^K, N_b^K - 1) \geq 0 \end{aligned} \quad (3)$$

Using the same logic, we construct many “offsetting” swaps of stores between the three retailers.¹³ By considering many such “minor perturbations” to the observed spatial net-

¹²It would also have been possible to apply this differencing procedure to inequalities involving a single location, rather than pairs of locations. For example, instead of creating an alternative configuration with one less Target store in location a and one more Target store in location b , we might have considered an alternative with simply one less Target store in location a . This alternative would also be profit-reducing. Then, an offsetting inequality could be generated by adding one Kmart store to location a (also profit-reducing for Kmart). When combining the inequalities, the common θ_a would drop out. The current approach of double-differencing has the (small) advantage of holding the total number of stores fixed. This is important in that we do not attempt to model (nor does our approach require) the firm’s decision about how many stores to build in total.

¹³In particular, we randomly select 15,000 pairs of locations (with replacement) from the full set of existing stores. This effectively weights the draws by population, since more populated areas will have more stores. In addition, because there are twice as many Wal-Mart stores as Kmart or Target stores, a higher percentage of the inequalities will involve Wal-Mart-Kmart or Wal-Mart-Target pairings, than will involve Target-Kmart pairings. The choice of 15,000 itself is somewhat arbitrary; similar results were obtained using fewer and using more random pairs.

work of stores, we are able to construct a pairwise maximum score objective function with which to estimate the β parameters. (The remaining (common) parameters, γ^X , are estimated in a second step.) Because we have explicitly controlled for the structural error in ξ_j , we now assume that any remaining idiosyncratic error leading to violations of inequality (3) is median independent of the included covariates. Continuing with the example of our Target-Kmart swap:

$$\begin{aligned} \widetilde{\Delta V}_a^T (N_a^T, N_a^T - 1) + \widetilde{\Delta V}_b^T (N_b^T, N_b^T + 1) \\ + \widetilde{\Delta V}_a^K (N_a^K, N_a^K + 1) + \widetilde{\Delta V}_b^K (N_b^K, N_b^K - 1) \geq v_{ab}^{TK} \end{aligned} \quad (4)$$

The error v_{ab}^{TK} rationalizes the fact that the inequality will not necessarily hold for every store swap that we consider. As noted above, we assume that this disturbance captures expectational (forecast) error in the agent’s beliefs over the conditioning variables that is median independent of their recorded values.¹⁴ Under this assumption, the rank-ordering property defined by Manski (1975) and further described by Fox (2007) is preserved, and consistent estimates of the parameters can be obtained using only this subset of the almost uncountable number of alternative configurations. Manski (1985) established that point identification of the β parameters requires at least one regressor with absolutely continuous support on the real line; a regressor with dense support in critical regions is needed to “break ties” between candidate β ’s, thereby winnowing the identified set to a point. Distances to the nearest distribution center and the firm’s headquarters, which will always vary across locations and players, are natural candidates for such a regressor since we only consider switching stores from one market to another. In the absence of this special regressor, maximum score could still be used to trace out the identified set, but a different inference procedure would be required to construct confidence sets.

¹⁴In the notation of Pakes (2010), $v_{1,i}$ plays a role analogous to our $v_{ab}^{f_1 f_2}$. Pakes describes this error as being composed of two terms. The first represents “the difference between the agent’s expectation of profits at the time the agent makes its decision and the realization of profits. It is a result of uncertainty in the exogenous variables that will eventually help determine profits.” He describes the second component as “that part of the error in our measure of profits that is mean independent of the information the agent bases its decision on. It will be generated by specification error in our profit functions and/or measurement error in the variables used to construct profits.” The non-linearity of the maximum score objective function invalidates the second interpretation in our context. However, as a check on our method, we estimated the parameters of the model using a moment inequality approach instead (i.e. minimizing the squared violations) and found similar results, hopefully mitigating these concerns.

In practice, the non-smoothness of the maximum score objective function makes optimization difficult, so we use smoothed maximum score with a normal kernel, and bootstrap with sub-sampling to obtain standard errors.¹⁵ Since scale is arbitrary in any discrete choice model, we normalize the parameter vector by setting $\|\beta\| = 1$, where $\|\cdot\|$ is the standard L^2 (Euclidean) norm. All of the parameter estimates are therefore expressed in this metric.

Finally, it is important to note what has happened with the endogenously determined local attributes $(N_j^f, N_j^{-f}, DC_j^f)$ in this system. We would expect each of these to be correlated with ξ_j . That is, places that are desirable owing to unobserved factors are likely to have more stores in them, both own and rival. Locations surrounded by locations with desirable unobservables are, similarly, more likely to be close to a distribution center (assuming distribution centers are placed with the goal of serving a large number of attractive markets). The key to our estimation strategy is that ξ_j no longer appears in our objective function when it comes time to estimate the parameter vector β . However, the other elements of θ_j (apart from ξ_j), which are clearly critical to giving a meaningful economic interpretation to several of the parameters recovered in the first stage, have also been differenced away. This is an issue in any procedure that involves differencing out common unobservables. An important contribution of our framework is a method for recovering these additional parameters in a second stage. We turn to this exercise next.

3.2 Decomposing the Market-Level Fixed Effects

Crucial to our understanding of firm entry behavior is the recovery of the parameters γ^X , which are included in θ_j , the fixed effect terms that were differenced out of equation (3). Recall from above that these parameters describe how the demographic variables included in X_j affect Wal-Mart's profits. The impact of X_j on Kmart's and Target's profits are represented by $f(X_j; \gamma^X) + \beta^{K,X} X_j$ and $f(X_j; \gamma^X) + \beta^{T,X} X_j$, respectively. Although we have already obtained estimates of $\beta^{K,X}$ and $\beta^{T,X}$ in the first stage, we clearly need to know Wal-Mart's values (captured only by $f(X_j; \gamma^X)$) in order to determine the overall value that any firm places on an attribute like per-capita income or population. Because we were concerned that ξ_j would be correlated with all the N 's, we included it in the fixed effect

¹⁵Formal proofs of consistency are provided in Fox (2007) and Bajari and Fox (2013), which build on arguments developed in Han (1987).

(θ_j) and differenced it out in the first stage. A key advantage of our two-stage approach is that estimates of the first-stage parameters can be recovered without having to specify a particular functional form for $f(X_j; \gamma^X)$. However, if we want to learn the baseline effects of X_j on all firms' profitabilities, we need to commit to a particular functional form; here, we use $\gamma^X X_j$. We can then use our first stage estimates, along with an assumption on firm behavior, to recover (set valued) estimates of the fixed effects. We can then project these intervals onto X using Beresteanu and Molinari's (2008) techniques for set valued random variables.

We begin this process by recovering estimates for the market-specific fixed effects. With only three firms and without making explicit distributional assumptions, we are unable to recover point estimates of each θ_j . There is simply not enough information in data describing the store siting decisions of just three firms to identify the precise values of this many parameters. However, we can recover ranges in which those parameters must lie and use these intervals to set-identify γ^X . In order to do so, we make use of an additional assumption about firm behavior similar to that employed by Pakes et al. (2011). In particular, we begin with each firm's marginal profitability of stores in each market j :

$$\frac{\partial \Pi^f}{\partial N_j^f} = \frac{N_j^f}{N_j^f + 1} \left[\beta^{f,Own} + 2\beta^{f,Own2} \ln(N_j^f + 1) + \beta^{f,Own3} \ln(Pop_j) \right] + \pi_j^f$$

Note that we will not attempt to recover the determinants of the total number of stores that each firm builds. This will be affected by capital constraints and cash reserves, long-run business plans, access to foreign suppliers, and so forth. Instead, we assume only that each firm allocated whatever stores it did build in an optimal fashion – that is, they added stores to each market until the marginal profitability of the last store added was equalized across markets, up to an integer constraint. In particular, for each firm f , there is a constant value c^f such that observed marginal profits in each market with at least one store must be greater than c^f , while the marginal profits from one additional store must be less than c^f :

$$\frac{N_j^f}{N_j^f + 1} \left[\beta^{f,Own} + 2\beta^{f,Own2} \ln(N_j^f + 1) + \beta^{f,Own3} \ln(Pop_j) \right] + V_j^f(N_j^f) + \theta_j \geq c^f \quad (5)$$

$$\frac{N_j^f + 1}{N_j^f + 2} \left[\beta^{f,Own} + 2\beta^{f,Own2} \ln(N_j^f + 2) + \beta^{f,Own3} \ln(Pop_j) \right] + V_j^f(N_j^f + 1) + \theta_j < c^f. \quad (6)$$

Note that in markets where the firm has zero stores, it is at a corner solution and only equation (6) must hold. We define the following components of firm f 's marginal profits in market j :

$$\begin{aligned} \psi_{1,j}^f &\equiv \frac{N_j^f}{N_j^f + 1} \left[\beta^{f,Own} + 2\beta^{f,Own2} \ln(N_j^f + 1) + \beta^{f,Own3} \ln(Pop_j) \right] + V_j^f(N_j^f) \\ \psi_{2,j}^f &\equiv \frac{N_j^f + 1}{N_j^f + 2} \left[\beta^{f,Own} + 2\beta^{f,Own2} \ln(N_j^f + 2) + \beta^{f,Own3} \ln(Pop_j) \right] + V_j^f(N_j^f + 1) \end{aligned}$$

Since we have estimated all the β parameters in the first stage (and observe everything else), we can treat this component of marginal profits as known.¹⁶ Because payoffs are only identified up to an additive scale parameter in any discrete choice framework, we have one free normalization. Setting the value of θ_j equal to zero for some base location, we have:¹⁷

$$\psi_{2,base}^f < c^f \leq \psi_{1,base}^f$$

allowing us to “back out” bounds on c^f from the profitability conditions.

With these bounds on c^f for each firm, we can now use equations (5) and (6) to bound θ_j for each firm:

$$\begin{aligned} \psi_{2,base}^f - \psi_{1,j}^f < \theta_j &< \psi_{1,base}^f - \psi_{2,j}^f, & \text{if } N_j^f > 0 \\ \theta_j &< \psi_{1,base}^f - \psi_{2,j}^f, & \text{if } N_j^f = 0 \end{aligned}$$

Assuming that the θ_j 's are the same for all firms, the intersection of these bounds

¹⁶We correct for the fact that these are estimates in the second stage standard errors by bootstrapping over the entire procedure.

¹⁷In practice, the base location used is Visalia, California, which has four Wal-Marts, three Kmart and two Targets.

describes the set of admissible values that θ_j can take for each market j .¹⁸ Note that if we had point estimates of the θ_j 's, we could simply project them onto X by linear regression. However, because the θ_j 's are set identified, our dependent variable now comes in interval form: $\theta_j \in [\theta_{Lj}, \theta_{Uj}]$. Therefore, we rely on the methods developed by Beresteanu and Molinari (2008), which use a transformed Minkowski average of the data to recover bounds and confidence sets for γ^X . This is described in detail in Appendix B.

4 Results

Our main empirical results are presented in Tables 1 and 2 for several alternative specifications of the model. Table 1 reports the parameter estimates for the firm-specific components of per-store profits, which were recovered in the first stage of our two-step procedure. The results from the second stage are presented in Table 2. Column I of Table 1 contains the results from our preferred specification, which allows for heterogeneous rival effects that depend on both the firm in question and on the rival's identity. Column II reports results from a more parsimonious specification that restricts rival effects to be the same regardless of the particular rival's identity. Column III adds market-level median income to the first stage of the specification used in Column II, while Column IV removes the location-specific unobservables (ξ 's) from that specification. Finally, Column V presents a variant of the specification used in Column II that uses only recent store placements by Target and Wal-Mart, as a robustness check.

Focusing on the results from the first stage of our preferred specification (Column I of Table 1), several features are worth noting. Since this is a highly non-linear specification, we begin with a simple discussion of signs and significance and then present marginal effects for the main variables of interest. The signs of all coefficients are as expected and most are significant at the usual levels. In particular, we find that rival stores have a strong negative and statistically significant impact on own profitability that is similar across firms and rivals. There are, however, a few small differences. Target and Kmart both suffer more from the entry of the other than they do from the entry of Wal-Mart, suggesting

¹⁸Note that we drop from the θ_j decomposition exercise those locations where the sets of θ_j 's are disjoint (i.e., where no value of θ_j can satisfy all three firms inequalities). This is an indication of model misspecification. Because this only happens with a small fraction of our overall locations (< 4%), we ignore any possible resulting sample selection in the decomposition of θ_j .

that they are the closer substitutes. Wal-Mart, on the other hand, suffers equally from the entry of either competitor. Note that the validity of this and other asymmetric rivalry effects depend crucially on the twin assumptions that 1) we have adequately controlled for all *observed* determinants of profits that different firms might value differently, and 2) any remaining *unobserved* components are indeed valued equally by all three players.

Agglomeration effects are described by the three parameters for each firm corresponding to terms containing N_{own} . When it comes to their own stores, Target and Wal-Mart show evidence of greater agglomeration effects than Kmart, but these effects fall off more quickly with rising population. Turning to the remaining parameters, all firms show evidence of a negative effect from increasing distance to both headquarters and distribution centers. The estimates also suggest that Target values large population centers more than Wal-Mart (although this effect is statistically insignificant), while Kmart values them substantially less. Finally, the smoothed maximum score as a fraction of the total number of comparisons can be used as a measure of fit. The value here is 14546 or 97%, indicating a very tight fit (note the sharp contrast with the specification in Column IV that does not include the market-level unobservable ξ).¹⁹

The specification in Column II restricts the “other” effect to be the same across rivals; we impose this restriction in order to demonstrate the implications of some alternative assumptions in the remaining columns. Under this parsimonious specification, “own” and “rival” effects become more heterogeneous. Target’s relatively stronger preference for population becomes statistically significant (while Kmart’s relatively weaker preference for population remains so), the adverse consequences of distance to headquarters lose significance, but the detrimental effects of distance to distribution center become even stronger.

The specification in Column III allows Target and Kmart to have preferences for market-level log median income that differ from those of Wal-Mart (whose preferences are captured in the baseline). Both Kmart and Target indeed show a greater preference for higher income locations relative to Wal-Mart, but neither of these differences is statistically significant.

¹⁹Population alone captures many of the relevant differences in demand across markets, while the common unobservables (the ξ_j 's) account for the bulk of the rest. Other specifications not reported here include measures of population density, average household size, the number of other supercenters (e.g. Meijer, Fred Meyer, Fry's), and the number of smaller retailers. The results for the competitive and own-firm effects showed a similar pattern, but the coefficients on the additional demand characteristics were insignificant. Specifications that included only the larger markets also yielded similar results.

The remaining parameters in Column III are similar to those in Column II. We return to the remaining specifications in Columns IV and V below.

Table 1: First Stage Estimates

	I		II		III		IV		V	
	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.
Kmart-Specific Covariates										
$\ln(N_{own} + 1)$	-.0808	.0126	-.339	.023	-.348	.023	-.211	.267	§	
$(\ln(N_{own} + 1))^2$.0002	.0067	-.038	.0027	-.043	.0026	.031	.097		
$\ln(Pop)(\ln(N_{own} + 1))$	-.0179	.0026	.0299	.0021	.0317	.0020	.005	.096		
$\ln(N_{Target} + 1)$	-.4079	.0098								
$\ln(N_{Walmart} + 1)$	-.3792	.0094								
$\ln(N_{Other} + 1)$			-.620	.011	-.6214	.0134	-.434	.059		
$\ln(\text{Distance to HQ})$	-.0013	.0006	-.0039	.0025	-.0038	.0022	-.072	.061		
$\ln(\text{Population})$	-.050	.010	-.080	.014	-.114	.016	.310	.091		
$\ln(\text{Income})$.022	.017				
Target-Specific Covariates										
$\ln(N_{own} + 1)$	-.046	.0067	-.134	.025	-.140	.0178	.085	.245	-.046	.016
$(\ln(N_{own} + 1))^2$.014	.0058	.0048	.0036	.0084	.0032	.082	.166	.0038	.002
$\ln(Pop)(\ln(N_{own} + 1))$	-.0288	.0028	-.0084	.0029	-.010	.0023	-.041	.158	-.0179	.002
$\ln(N_{Kmart} + 1)$	-.423	.0076								
$\ln(N_{Walmart} + 1)$	-.3798	.0105								
$\ln(N_{Other} + 1)$			-.527	.011	-.518	.010	-.276	.199	-.596	.0035
$\ln(\text{Distance to HQ})$	-.0153	.0077	-.0012	.0041	.0012	.0037	-.071	.114	.0033	.0015
$\ln(\text{Population})$.0145	.0107	.042	.016	.013	.016	.436	.156	.021	.013
$\ln(\text{Income})$.012	.019				
Wal-Mart-Specific Covariates										
$\ln(N_{own} + 1)$	-.039	.012	-.240	.024	-.208	.022	-.295	.129	-.052	.015
$(\ln(N_{own} + 1))^2$.0141	.0072	.0095	.0041	.015	.0036	.102	.029	.008	.003
$\ln(Pop)(\ln(N_{own} + 1))$	-.0267	.0032	-.011	.003	-.017	.0029	-.028	.017	-.027	.002
$\ln(N_{Target} + 1)$	-.4146	.012								
$\ln(N_{Kmart} + 1)$	-.4235	.0074								
$\ln(N_{Other} + 1)$			-.369	.010	-.381	.0089	-.376	.063	-.438	.005
$\ln(\text{Distance to HQ})$	-.0010	.0002	.0019	.0056	-.0028	.0050	-.090	.049	.025	.004
$\ln(\text{Population})$			‡		‡		.358	.052	‡	
Common Covariates										
$\ln(\text{Distance to DC})$	-.00101	.00015	-.0032	.0010	-.0050	.0010	-.029	.012	-.0031	.0011
Number of Comparisons	15000		15000		15000		30000		15000	
Smoothed Maximum Score	14546		13649.2		13657.5		21657		13193.43	

‡ These common components of the profit function are differenced out as part of the location fixed effect. They are estimated in a separate second stage, the results of which are reported in Table 2. In Specification IV, because we have dropped the unobservable location-specific characteristic, there is no need to use the differencing approach; all parameters can be recovered in a single stage.

§ Specification V estimates the model using only stores placed since 2000 to construct inequalities. Since Kmart did not open any new stores between 2000 and 2006, we do not estimate coefficients for this firm.

The results from the second stage are reported in Table 2. The bounds and 95% confidence sets for the coefficient on log population are positive in all our specifications, indicating, as expected, that larger markets are more profitable.²⁰ We also note that, while

²⁰Note that the bounds recovered from the set-inference procedure are conservative in that they represent

Table 2: Second Stage Estimates

	I	III	V
ln(Population)	(.3179,.4095) [.2883,.4392]	(.2212,.2888) [.1996,.3104]	(.2481,.3496) [.2344,.3633]
ln(Median Income)	(-.3555,.3221) [-.4189,-.3848]	(-.3034,.2327) [-.3335,.2628]	(-.3139,.4369) [-.3426,-.4656]
Constant	(-.0032,-.00054) [-.00354,-.000176]	(-.00227,-.00017) [-.00243,-.000014]	(-.0033,-.00037) [-.00347,-.000205]

Each column uses the first-stage estimates from specifications I, III, and V (respectively) in Table 1. The intervals appearing in square brackets represent the largest 95% confidence collection, i.e., the largest set that, when specified as a null hypothesis for the true value of the population identification region, cannot be rejected by a 95% test.

the first-stage estimates indicate Wal-Mart’s profitability from population is lower than Target’s, the net effect (captured by γ^{pop}) is still positive. The same is true for Kmart, which places an even lower premium on larger population. The set estimates for log median income are neutral; that is, they cover both positive and negative values. It is therefore difficult to draw conclusions about these firms’ preferences for income, after controlling for the impact of high values of population. This is likely due to the relatively coarse market definition we have employed (within-market comparisons might yield greater variation). However, pinning down the baseline impact of population does allow us to quantify the impact of own and rival stores in terms of required changes in population (the main driver of demand).

Table 3 summarizes the impact of “own” and “rival” effects on profitability. There, we present the increase in population required to offset the entry of a rival store or induce the entry of another own store, evaluated for a few representative market configurations using the midpoint of the second stage population coefficient. For the smallest market category, an additional 36,000 to 41,000 people are needed to induce these firms to build another store, about 50% larger than the “back of the envelope” minimum efficient scale calculation discussed in Section 2 and consistent with a strong initial own-competition effect. Note that this “required” level of population increases with market size, as discount stores likely face the outer envelope of values that the parameter can take for all possible values of the remaining parameters. Our results therefore imply that, *no matter what value is taken by the vector of other parameters*, log population always has a positive effect on profits. The same cannot be said for income.

more and better substitutes in larger markets. Moreover, this effect increases quite sharply at first and then levels off, perhaps reflecting the fact that the urban amenities produced by a “city” are quite similar above a minimum population threshold of 200,000 or so. This seems consistent with casual empiricism.

Table 3: Increase in Population Required to Offset or Induce Entry

Initial # of Own Stores, Population	Change	Kmart	Target	Walmart
Originally 1 own store, population of 50,000	1 → 2 own stores	41,078	38,134	36,132
	1 → 2 rivals	68,011	49,419	53,349
Originally 2 own stores, population of 150,000	2 → 3 own stores	89,718	83,881	79,483
	2 → 3 rivals	132,531	100,194	107,370
Originally 4 own stores, population of 250,000	4 → 5 own stores	93,952	86,846	82,170
	4 → 5 rivals	131,287	103,213	109,828
Originally 8 own stores, population of 500,000	8 → 9 own stores	112,231	102,908	97,080
	8 → 9 rivals	148,083	120,623	127,852

Looking across the three firms and across market sizes, Wal-Mart requires the smallest increase in population to offset another of its own stores. Target, on the other hand, requires the smallest increase in population to offset the entry of a rival store, perhaps reflecting its relatively unique positioning amongst the three. Kmart is hurt the most by either form of entry, which is not surprising given its overall market performance.

One mechanism that could explain these results is that Wal-Mart’s efficient distribution network better capitalizes on proximate stores. Wal-Mart may also be better equipped to share advertising and knowledge about local markets across its nearby outlets. Both are strengths that have been noted in the trade press. It is also possible that Target’s customers are more brand-loyal, making it less likely to lose them to a new rival entrant. Kmart, which filed for bankruptcy in 2002, exhibits none of these strengths.

4.1 The Role of Unobserved Location Attributes

One of the key advantages of our modeling strategy is the ability to account for spatial unobservables, which we denote by ξ_j , that might be important determinants of firms’ siting decisions. If firms site more stores in locations with attractive unobservables, but those unobservables are ignored in estimation, the researcher could mistake them for positive spatial spillovers (or will at least understate the role of business stealing effects – both between a

firm's own stores and with those of its rivals). In order to illustrate the important role played by these unobservables, we re-estimate the model used above under the restriction that $\xi_j = 0 \forall j$. Estimation is then based on the following inequality

$$N_a^f \cdot \hat{\pi}_a^f(N_a^f) + N_b^f \cdot \hat{\pi}_b^f(N_b^f) \geq (N_a^f - 1) \cdot \hat{\pi}_a^f(N_a^f - 1) + (N_b^f + 1) \cdot \hat{\pi}_b^f(N_b^f + 1)$$

where

$$\begin{aligned} \hat{\pi}_j^f(N_j^f) = & \beta^{f,Own} \ln(N_j^f + 1) + \beta^{f,Own2} (\ln(N_j^f + 1))^2 + \beta^{f,Own3} (\ln(Pop_j)) (\ln(N_j^f + 1)) \\ & + \beta^{f,Other} \ln(N_j^{-f} + 1) + \beta^{f,HQ} \ln(HQ_j^f) + \beta^{f,DC} \ln(DC_j^f) + \alpha^{f,X} X_j. \end{aligned}$$

The main difference between these equations and those employed under the full (differenced) specification presented earlier is that

$$\beta^{f,X} X_j + \theta_j \quad (f = \text{Kmart, Target, Wal-Mart})$$

is replaced by

$$\alpha^{f,X} X_j \quad (f = \text{Kmart, Target, Wal-Mart}).$$

Without θ_j , all parameters can therefore be estimated in a single-stage, using the maximum score procedure detailed above. In particular, we estimate a separate log population parameter for each firm in the first stage. Because log income was not significant in the second stage of our other specifications, we omit it here.

Column IV of Table 1 reports these results. It is easy to see from the parameter estimates that the effects of competition from rivals are all biased towards zero. Rival stores are clearly proxying for unobserved spikes in the relative desirability of particular markets. The coefficient on the distance to the nearest distribution center is biased away from zero, implying that distance to DC and desirable unobserved characteristics are negatively correlated. This suggests that firms site their distribution centers closer to the unobservably 'good' places to locate a store, an intuitive result. Without ξ , the model confounds unobserved features of the market that are attractive to all three firms with a desire to cluster their stores. A similar effect can be seen by looking at the coefficients on the N_{own} terms – without the unobserved location attribute, the model tends to over-estimate the benefits of agglomeration.

Finally, the importance of unobservable characteristics can also be seen by comparing the value of the score function under different specifications. When the unobserved market characteristic is ignored (Column IV), the estimator correctly satisfies approximately 72% of the sampled inequalities. This score jumps to 97% when the unobservable is accounted for, highlighting the role of common market-level unobservables in explaining entry behavior.

4.2 Robustness Check: Estimation Based on Only Recent Store Placement

While the inherent non-stationarity of the discount store industry makes a fully dynamic model of retail competition intractable using current methods, we can examine the robustness of our static framework by focusing on recently placed stores, taking the pre-existing locations as given. In particular, we re-estimate the model using only those stores sited after 2000 (the earliest date for which we could obtain data from *Trade Dimensions*). As noted earlier, Kmart has not opened any stores in the past decade, so this part of the analysis will focus only on the actions of Wal-Mart and Target alone (although we do condition on the presence of existing Kmart stores). Both Wal-Mart and Target were quite active during this period (2000-2006), with Wal-Mart adding 760 new stores in 294 markets and Target adding 471 new stores in 168 markets.

The estimates from this specification are reported in Column V of Table 1. They share similarities with both Columns I and II, providing a degree of validation and assurance that our results are not being driven by stores that were sited long ago. Own store effects are similar to Column I, while rival store effects reflect a mix of the results in Columns I and II. The log of distance to the nearest distribution center is still significantly negative, and similar in magnitude to Column II. Target's higher relative preference for log population continues to be significant as well. Finally, the coefficient on distance to headquarters is now positive and significant for both Target and Wal-Mart, reflecting the fact that the newest stores are mostly farther from each firm's home base, as the closer-in locations were added earlier on.

4.3 Counterfactual: Illustrating the Role of Positive Spillovers within a Chain

The previous section characterized the relative magnitude of rival and own (chain) effects using comparative statics. We now address the full equilibrium implications of the local chain effect through a counterfactual exercise. In particular, we use the first order conditions described in Section 3.2 to solve for the new configuration that obtains when local chain effects are removed. This exercise highlights two important sources of flexibility in our modeling strategy. First, we can accommodate any number of firms, while previous approaches were restricted to either a single firm (Holmes (2011)) or a pair of firms (Jia (2008), Nishida (2008)). Second, since our model does not require that spatial spillovers be positive, we can handle cases (such as this) in which business stealing overwhelms local chain effects on net. To capture the equilibrium impact of local chain effects, we construct a counterfactual scenario in which the three firm’s treat each of their outlets as a single profit center, rather than maximizing profits for the full set taken together. In particular, we replace the earlier per store profit function with the following

$$\begin{aligned} \pi_j^f = & \beta^{f,Own} \ln(2) + \beta^{f,Own2} (\ln(2))^2 + \beta^{f,Own3} (\ln(Pop_j)) (\ln(2)) + \beta^{f,Other1} \ln(N_j^f + N_j^{-f1}) \\ & + \beta^{f,Other2} \ln(N_j^{-f2} + 1) + \beta^{f,HQ} \ln(HQ_j^f) + \beta^{f,DC} \ln(DC_j^f) + \beta^{f,X} X_j + \theta_j \end{aligned}$$

Consequently, apart from the current store under consideration, the firm treats its own remaining stores symmetrically to those of its least damaging rival (“rival 1” in this notation). Under this counterfactual, firms no longer benefit from the density economies (and other chain synergies) that cause the net impact of one’s own stores to be smaller than the competition effect from rivals. This will impact firms heterogeneously depending on the degree to which they were successful in exploiting these economies under the original scenario, yielding insight into each firm’s differential comparative advantage. Firms will therefore no longer fully internalize the externality caused by their own local stores, which presumably limits their incentives to build additional stores in practice.

To simulate this counterfactual scenario, we re-solve for the equilibrium number of stores operated by each firm using the estimated parameter values from Column I of Table 1 (heterogeneous rival effects) and the profit function specified above, using iterated best

response. This task is simplified considerably by the fact that, taking the location of distribution centers as given, all spillovers occur within market, thus allowing us to re-solve for the equilibrium on a market by market basis.²¹ Recall that, because there are only three players, we could only bound the values of c^f and θ_j in estimation.²² For purposes of the counterfactual, we use the midpoints of these intervals as point estimates of these parameter values.

Table 4 summarizes the results of this counterfactual simulation. Not surprisingly, in most of the small metro and non-metro markets, the equilibrium configuration does not change under the counterfactual since many of these markets were served by at most one store of each type to begin with (often just a single store). The counterfactual scenario should not have much impact on these markets. As the markets increase in size, the change increasingly kicks in, impacting the three firms quite differently. Wal-Mart, which benefits most from local chain effects, is hurt the most under this scenario, and ends up opening 790 fewer stores (mostly in medium and large metro markets). In contrast, Target responds to this reduction in competition (and the fact that it is hurt the least by rival stores) by adding 1115 outlets, mostly in the largest markets where it already enjoyed a comparative advantage. The increased competition from Target and the absence of any local chain economies results in Kmart opening fewer stores on net (-147). The change is strongest in the largest markets, which Kmart and Wal-Mart essentially concede to Target. However, Kmart does manage to add stores in the smallest non-metro markets, many of which are closer to their distribution network than Wal-Mart’s more diffuse layout.

This counterfactual exercise provides an insight into an important force behind Wal-Mart’s role as leader in the “big-box” chain store industry. In particular, Wal-Mart goes

²¹There are multiple solutions to the system of best response equations, reflecting multiplicity in the underlying discrete game. We “solve” this multiplicity problem by selecting the equilibrium that is closest to the prior market configuration (i.e. the one in the data). Alternative selection rules yielded qualitatively similar results.

²²Here we have held the marginal profitability of the last store, c^f constant for each firm. In essence, this assumes that the firm’s access to capital remains fixed in the counterfactual scenario. Because stores themselves become more or less profitable in the counterfactual scenario, the total number of stores that are built will change. An alternative approach would have been to hold the total number of stores owned by each firm fixed, and allow them to allocate those stores so that the marginal profitability of the last store is the same across locations, but takes on a different value than the c^f we observe. Without a specific model for how firms construct their distribution networks, we cannot predict how these networks might change under different counterfactual scenarios.

Table 4: Counterfactual: Change in Number of Stores

Type of Market	Average Population	Markets with No Change	Average & Net Change in			
			Total Stores	Kmart	Target	Wal-Mart
Isolated Counties	24,564	82.9%	.10	.08	.86	-.11
Small Micro/Metropolitan Area	71,677	75.9%	47	36	61	-50
			-.172	-.02	.14	-.23
Metro > 200,000	323,555	4.9%	-123	-1	45	-167
			-.57	-.46	.06	-2.54
Metro > 500,000	691,872	0%	-59	-48	251	-262
			6.8	-2.9	2.44	-6.76
All	97,649	69.9%	313	-134	758	-311
			.14	-.11	16.47	-.61
			178	-147	1115	-790

from having 54% of all stores (compared to Target’s 23%) in reality, to being evenly matched with Target at 40% in the counterfactual without local chain effects. Without detailed data on prices and costs, we cannot say anything definitive about consumer surplus, but it seems likely that this more competitive structure would yield tougher price competition and lower prices.

5 Conclusions

Many of the most important sectors in the U.S. economy are characterized by a small number of firms making many interlinked entry decisions over a large choice set (i.e., a high-dimension network game). This market structure creates considerable difficulty for the economic modeling of entry behavior. In particular, the large number of possible configurations available to firms rules out using the traditional likelihood-based approaches (assuming either complete or incomplete information) to modeling games. In this paper, we propose an alternative approach based on profit inequalities (Pakes et al. (2011), Bajari et al. (2007)). Basing identification only on the premise that the firm chose its best possible configuration (given the equilibrium play of its rivals), this approach avoids ever having to solve for an equilibrium or even specify a selection mechanism. Nor does it ever require that we solve for the probability of seeing each of an almost uncountable number of configurations. Implementing this estimator is complicated by the fact that we need to control for the effect on entry of market attributes, many of which are unobserved. We show how this can be accomplished by using the logic of a difference-in-differences estimator nested inside these inequalities.

We conclude by noting two limitations of our analysis (or, alternatively, opportunities for future research). First, we model entry behavior as a static game. Even with the methods described above, this is a challenging exercise, but it clearly ignores the dynamics that unfolded over the long time frame over which Target, Kmart, and Wal-Mart built their stores. Adding dynamics while maintaining our full strategic environment, focus on distribution network, and choice set definition poses serious challenges. Second, we are forced to restrict our structural errors to only allow for a market-level unobservable that is common across all firms. While this omission would likely bias our estimates in the direction of finding strong local chain effects (strengthening our empirical results), it would be desirable to allow firms to have heterogeneous preferences for market features that are unobserved by the econometrician. Doing so in the present context would, however, invalidate our “double difference” approach to handling market-level unobservables.

References

- [1] **Akerberg, D. and G. Gowrisankaran.** (2007). “Quantifying Equilibrium Network Externalities in the ACH Banking Industry,” *RAND Journal of Economics*, 37(3): 738-761.
- [2] **Andrews, D., and P. Jia.** (2012). “Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure,” *Econometrica*, 80(6): 2805-2826.
- [3] **Bajari, P., L. Benkard, and J. Levin** (2007). “Estimating Dynamic Models of Imperfect Competition,” *Econometrica*, 75(5): 1331-1370.
- [4] **Bajari, P. and J. Fox** (2013). “Measuring the Efficiency of an FCC Spectrum Auction,” *American Economic Journal: Microeconomics*, 5(1): 100-146.
- [5] **Bajari, P., H. Hong, J. Krainer, and D. Nekipelov** (2010). “Estimating Static Models of Strategic Interactions,” *Journal of Business and Economic Statistics*, 28(4): 469-482.
- [6] **Bajari, P., H. Hong, and S. Ryan** (2010). “Identification and Estimation of a Discrete Game of Complete Information,” *Econometrica*, 78(5): 1529-1568.
- [7] **Basker, E.** (2005). “Job creation or destruction? Labor market effects of Wal-Mart expansion,” *Review of Economics and Statistics*, 87: 174–183.
- [8] **Basker, E.** (2007). “The Causes and Consequences of Wal-Mart’s Growth,” *Journal of Economic Perspectives*, 21(3): 177-198.
- [9] **Beresteanu, A. and P.B. Ellickson** (2006). “The Dynamics of Retail Oligopoly,” Working Paper: University of Rochester.
- [10] **Beresteanu, A., I. Molchanov, and F. Molinari** (2011). “Sharp Identification Regions in Models with Convex Predictions: Games, Individual Choice, and Incomplete Data,” *Econometrica*, 79(6): 1785-1822.

- [11] **Beresteanu, A. and F. Molinari** (2008). “Asymptotic Properties for a Class of Partially Identified Models,” *Econometrica*, 76(4): 763-814.
- [12] **Berry, S.** (1992). “Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 60(4): 889-917.
- [13] **Bresnahan, T. and P. Reiss** (1990). “Entry in Monopoly Markets,” *Review of Economic Studies*, 57(4): 531-553.
- [14] **Bresnahan, T. and P. Reiss** (1991). “Entry and Competition in Concentrated Markets,” *Journal of Political Economy*, 99(5): 977-1009.
- [15] **Ciliberto, F. and E. Tamer** (2009). “Market Structure and Multiple Equilibria in Airline Markets,” *Econometrica*, 77(6): 1791-1828.
- [16] **Eizenberg, E.** (2012). “Upstream Innovation and Product Variety in the U.S. Home PC Market,” Working Paper: Hebrew University.
- [17] **Ellickson, P.B.** (2007). “Does Sutton Apply to Supermarkets?,” *RAND Journal of Economics*, 38(1): 43-59.
- [18] **Ellickson, P.B. and P.L.E. Grieco** (2013). “Wal-Mart and the Geography of Grocery Retailing,” *Journal of Urban Economics*, 75: 1-14.
- [19] **Ellickson, P.B. and S. Misra** (2011). “Estimating Discrete Games,” *Marketing Science*, 30(6): 997-1010.
- [20] **Ellickson, P.B., S. Misra and H. Nair** (2012). “Repositioning Dynamics and Pricing Strategy,” *Journal of Marketing Research*, 49(6): 750-772.
- [21] **Fox, J.** (2006). “Estimating Matching Games with Transfers,” Working Paper: University of Chicago.
- [22] **Fox, J.** (2007). “Semiparametric Estimation of Multinomial Discrete Choice Models Using a Subset of Choices,” *RAND Journal of Economics*, 38(4): 1002-1019.
- [23] **Han, A.** (1987). “Nonparametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator,” *Journal of Econometrics*, 35(2-3): 303-316.

- [24] **Hausman, J. and E. Leibtag** (2007). “Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart,” *Journal of Applied Econometrics*, 22(7): 1157-1177.
- [25] **Holmes, T.** (2011). “The Diffusion of Wal-Mart and Economies of Density,” *Econometrica*, 79(1): 253-302.
- [26] **Hyslop, D.R. and G.W. Imbens** (2001). “Bias From Classical and Other Forms of Measurement Error,” *Journal of Business and Economic Statistics*, 19(4): 475-481.
- [27] **Jia, P.** (2008). “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retail Industry,” *Econometrica*, 76(6): 1263-1316.
- [28] **Mahajan, A., A. Tarozzi, J. Yoong, and B. Blackburn** (2011) “Bednets, Information, and Malaria in Orissa,” Working Paper: Stanford University.
- [29] **Manski, C.** (1975). “Maximum Score Estimation of the Stochastic Utility Model of Choice,” *Journal of Econometrics*, 3(3): 205-228.
- [30] **Manski, C.** (1985). “Semiparametric Analysis of Discrete Response,” *Journal of Econometrics*, 27(3): 313-333.
- [31] **Manski, C. and E. Tamer** (2002). “Inference on Regressions with Interval Data on a Regressor or Outcome,” *Econometrica*, 70(2): 519-546.
- [32] **Mazzeo, M.** (2002). “Product Choice and Oligopoly Market Structure,” *RAND Journal of Economics*, 33(2): 221-242.
- [33] **Neumark, D., J. Zhang, and S. Ciccarella** (2008). “The Effects of Wal-Mart on Local Labor Markets,” *Journal of Urban Economics*, 63(3): 405-430.
- [34] **Nishida, M.** (2008). “Estimating a Model of Strategic Store-Network Choice,” Working Paper: Johns Hopkins University.
- [35] **Pakes, A.** (2010). “Alternative Models for Moment Inequalities,” *Econometrica*, 78(6): 1783-1822.

- [36] **Pakes, A., J. Porter, K. Ho, and J. Ishii** (2011). “Moment Inequalities and Their Application,” Working Paper: Harvard University.
- [37] **Seim, K.** (2006). “An Empirical Model of Firm Entry with Endogenous Product-Type Choices,” *RAND Journal of Economics*, 37(3): 619-640.
- [38] **Smith, H.** (2004). “Supermarket Choice and Supermarket Competition in Market Equilibrium,” *Review of Economic Studies*, 71(1): 235-265.
- [39] **Tamer, E.** (2003). “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *Review of Economic Studies*, 70(1): 147-165.
- [40] **Zhu, T., and V. Singh.** (2009). “Spatial Competition with Endogenous Location Choices - An Application to Discount Retailing,” *Quantitative Marketing and Economics*, 7(1): 1-35.
- [41] **Zhu, T., V. Singh, and M. Manuszak.** (2009). “Market Structure and Competition in the Retail Discount Industry,” *Journal of Marketing Research*, 46: 453-466.

A Additional Description of Data

The data for the discount store industry are taken primarily from the *Trade Dimensions* Retail Tenant Database for 2006. This proprietary dataset contains all 6,150 Wal-Mart, Kmart, and Target stores in operation in the continental United States as of August 2006. These include both pure discount stores that carry general merchandise and newer super-center formats that also carry full grocery lines.²³ Because of the additional difficulty in modeling a firm’s choice of store format, we do not distinguish between these two types of stores in our current application.

Stores are assigned to markets based on the county in which they are located. Where possible,²⁴ adjacent counties are grouped together into single markets according to the 2005 Census definitions for CBSAs. These statistical areas contain an average of 1.9 counties and include both a core urban area and any adjacent counties that are closely linked economically and socially. Metropolitan areas are those with a core urban area with a population of at least 50,000, and micropolitan areas are those with a core urban area with a population from 10,000 to 50,000.

We believe that using CBSA’s to define markets is more appropriate than using counties alone since they typically define a more natural “shopping area” by grouping interrelated counties together. The same criticism that applies to using large counties (e.g., Los Angeles County) to measure markets also applies here – all stores in the largest metro areas probably do not compete equally with one another. However, this broad market definition allows us to capture any positive supply-side spillovers that may potentially occur between stores of the same chain located in the same MSA.

We have collected demographic and economic data on each market from the U.S. Census Bureau. Population estimates at the CBSA and county levels are available from the Census Population Division for July 2006. Household median income and average household size are aggregated from the county level from the 2000 Census. We also include the percentage of the metro or micro area population that is located in an urban area, collected from the

²³We exclude Sam’s Club stores, owned by Wal-Mart, because these warehouse clubs do not compete directly with Kmart and Target stores. Furthermore, they are served by separate distribution facilities from those that serve Wal-Mart stores.

²⁴A total of 439 counties with discount stores are relatively isolated and are not included as part of any Census-defined CBSA. These counties are included as individual markets in our estimation.

website of the Missouri Census Data Center.

The Trade Dimensions database contains the locations of each of the 113 distribution centers that serve the stores in our analysis. Wal-Mart is vertically-integrated, owning all 70 distribution centers that supply its stores. Kmart and Target operate most of the distribution centers that supply their stores, 13 and 24 centers, respectively. However, a small number of their stores rely on third-party distributors. We include the locations of one Merchants Distributors center and five SuperValu distribution centers that serve Kmart and Target stores, respectively. Using the Haversine formula, we calculate the distance from the population-weighted centroid of each market to the street location of the closest distribution center for each firm.

Summary statistics for these markets are provided in Tables 5 and 6. Of the 1351 markets that contain at least one discount store, 358 are metro areas, 554 are micro areas, and 439 are counties outside a metro or micro area. Population estimates for 2006 range from 3,643 (in Norton, VA) to over 18.8 million (in New York-Northern New Jersey-Long Island). The number of discount stores per market ranges from 1 to 161 stores (in the Chicago market), but this distribution is highly skewed. The median number of stores per market is 1, and 75% of markets have 3 or fewer stores. Table 7 documents the expansion of Target and Wal-Mart between 2000 and 2006. Kmart did not add stores in any markets during this time period; instead it closed 710 stores following its bankruptcy in early 2002.

Table 5: Summary Statistics by Market

	Obs	Mean	Std. Dev.	Min	Max
Population in 2006	1351	213,424	860,138	3,643	18,818,536
Land area (in sq. miles)	1351	1,448.41	1,961.37	5.67	27,259.87
Population density	1351	119.23	199.60	1.76	2,797.94
Average household size	1351	2.53	0.18	2.18	3.75
Median income	1351	\$34,742	\$6,816	\$16,271	\$73,874
Percentage urban population	1351	0.52	0.22	0	1
Metropolitan Statistical Area	1351	0.26	0.44	0	1
Micropolitan Statistical Area	1351	0.41	0.49	0	1
Northeast	1351	0.08	0.28	0	1
Midwest	1351	0.32	0.46	0	1
South	1351	0.46	0.50	0	1
West	1351	0.14	0.34	0	1
Distance to closest Kmart DC (in miles)	1351	192.08	113.45	0.14	661.64
Distance to closest Target DC (in miles)	1351	144.20	96.81	0.70	682.44
Distance to closest Wal-Mart DC (in miles)	1351	90.53	62.76	0.75	507.60
Number of discount stores	1351	4.55	11.84	1	161
Number of Kmart stores	1351	1.01	2.89	0	43
Number of Target stores	1351	1.07	4.53	0	72
Number of Wal-Mart stores	1351	2.48	5.34	0	85

Table 6: Summary Statistics by Store for each Firm

	Kmart		Target		Wal-Mart	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Population in 2006	2,254,909	3,926,236	3,523,082	4,301,172	1,671,212	2,993,719
Land area (in sq. miles)	3515.56	3795.60	4922.78	4200.67	3417.85	3719.22
Population density	513.46	634.61	700.23	717.17	371.05	495.78
Average household size	2.56	0.20	2.62	0.20	2.57	0.19
Median income	\$41,640.98	\$7,578.02	\$45,519.59	\$7,475.11	\$39,731.92	\$8,038.29
Percentage urban population	0.76	0.20	0.86	0.13	0.70	0.23
Average Weekly Sales Volume	\$241,942	\$85,461	\$670,149	\$184,103	\$991,846	\$346,157
Metropolitan Statistical Area	0.78	0.42	0.95	0.21	0.68	0.47
Micropolitan Statistical Area	0.17	0.38	0.04	0.20	0.19	0.39
Northeast	0.18	0.38	0.12	0.33	0.11	0.32
Midwest	0.31	0.46	0.28	0.45	0.25	0.43
South	0.33	0.47	0.34	0.47	0.49	0.50
West	0.18	0.39	0.26	0.44	0.15	0.36
Distance to HQ (in miles)	713.24	567.09	913.92	428.08	695.32	368.72
Distance to closest own DC (in miles)	153.17	121.57	117.83	95.08	75.93	52.57
Number of Stores	1362		1443		3345	

Table 7: Comparing All Markets with Markets of Recently Placed Stores

Firm	Target	Walmart
# Stores	1443	3345
# Markets	381	1309
# Net New Stores	471	760
# Locations with New Stores	168	294
<i>All Markets where Firm is Present</i>		
Avg. Population	648,895	219,059
Percent Metro	83%	27%
Percent Micro	15%	41%
<i>Markets with Expansion Since 2000</i>		
Avg. Population	1,042,212	743,352
Percent Metro	91%	70%
Percent Micro	7%	17%

B Decomposing Market-Level Fixed Effects with Beresteanu and Molinari (2008)

After recovering interval estimates of market-level fixed effects using the bounding procedure described in Section 3.2, we estimate the parameters for the best linear predictor of θ in the interval $[\theta_L, \theta_U]$ conditional on X . Suppose X were to consist of two variables, x_1 and x_2 . Then the population set-valued best linear predictor is defined as

$$\Gamma = \left\{ \left[\begin{array}{c} \gamma_1 \\ \gamma_2 \end{array} \right] : \left[\begin{array}{c} \gamma_1 \\ \gamma_2 \end{array} \right] = \left[\begin{array}{cc} E(x_1^2) & E(x_1 x_2) \\ E(x_2 x_1) & E(x_2^2) \end{array} \right]^{-1} \left[\begin{array}{c} E(x_1 \theta) \\ E(x_2 \theta) \end{array} \right], \theta \in [\theta_L, \theta_U] \right\}$$

The estimate $\hat{\Gamma}$ is obtained by using the sample analogs of the above expectations. The third term in brackets can simply be estimated by $[\frac{1}{J} X'X]^{-1}$, which we will denote $\hat{\Sigma}^{-1}$. The last term in brackets can be written as $\mathbb{E}(G)$, where G is a set-valued random variable reflecting all possible values of $X'\theta$, given that θ is bounded by θ_L and θ_U :

$$G = \left\{ \left[\begin{array}{c} x_1 \theta \\ x_2 \theta \end{array} \right] : \theta \in [\theta_L, \theta_U] \right\}$$

The sample analog to this expectation is given by the Minkowski average $\bar{G}_J = \frac{1}{J} \bigoplus_{j=1}^J G_j$. That is, for each observation j , G_j is a line segment with endpoints $(x_{1,j}\theta_{Lj}, x_{2,j}\theta_{Lj})$ and

$(x_{1,j}\theta_{Uj}, x_{2,j}\theta_{Uj})$. The Minkowski sum of G_j , from $j = 1..J$, adds all these segments to form a many-sided polygon. Figure 4 illustrates the Minkowski sum of line segments corresponding to ranges of θ for two different markets (in our application, we would sum over all 1351 markets).

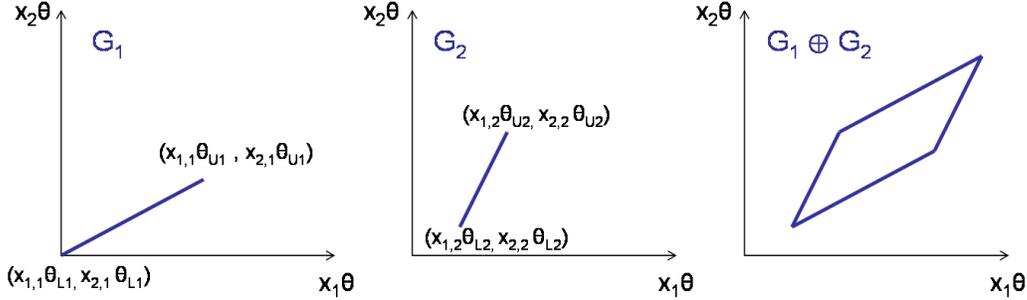


Figure 4: Minkowski Summation

The polygon generated by the Minkowski sum is then transformed by $(X'X)^{-1}$ to obtain an estimate of the set $\Gamma = [\gamma_1, \gamma_2]$. Therefore,

$$\hat{\Gamma} = \hat{\Sigma}^{-1} \bar{G}_J \quad (7)$$

In general, if there are K covariates in X , this estimated set $\hat{\Gamma}$ will be a K -dimensional polytope. In our application, the X variables include a constant term and various location-specific characteristics, such as the natural log of population and median income, the average household size, the percent of the area's population living in urban areas, the population density, and regional dummies. It is computationally difficult (if not impossible) to Minkowski sum such a high-dimensional set-valued random variable. However, we can estimate a subset of the parameters of the best linear predictor using the same logic as the Frisch-Waugh-Lovell Theorem for partial regression in point-identified models. For example, to estimate the set of possible coefficients on $\ln(\text{population})$ and $\ln(\text{median income})$, we first obtain the residuals from a linear regression of $\ln(\text{population})$ on all other X variables besides $\ln(\text{median income})$ and the residuals from a linear regression of $\ln(\text{median income})$

on all other X variables besides $\ln(\text{population})$. We then denote these \tilde{x}_1 and \tilde{x}_2 and use them in place of x_1 and x_2 above.

By the same partial regression logic, we can also estimate one-dimensional projections of the identification region. For example, the identification region of the best linear predictor coefficient for a single γ_k can be estimated by the interval ²⁵

$$\hat{\gamma}_k = \frac{1}{\frac{1}{J} \sum_{j=1}^J \tilde{x}_{k,j}^2} \left[\frac{1}{J} \sum_{j=1}^J \min \{ \tilde{x}_{k,j} \theta_{Lj}, \tilde{x}_{k,j} \theta_{Uj} \}, \frac{1}{J} \sum_{j=1}^J \max \{ \tilde{x}_{k,j} \theta_{Lj}, \tilde{x}_{k,j} \theta_{Uj} \} \right] \quad (8)$$

where $\tilde{x}_{k,j}$ is the residual from the j^{th} observation in a regression of x_k on the other x_{-k} , including the constant term.

For both the one-dimensional and two-dimensional projections, confidence sets can be formed by bootstrapping and computing the Hausdorff distance between the estimated set for the original sample and the estimated set for each bootstrapped sample, $H(\Gamma^b, \hat{\Gamma})$.²⁶ Under standard regularity conditions, Beresteanu and Molinari (2008) show that $r = \sqrt{n}H(\Gamma^b, \hat{\Gamma})$ is asymptotically normally distributed. Therefore, we can use the 95th percentile of the empirical distribution of r to construct bounds on the collection of all sets that, when specified as the null hypothesis for the true value of the population identification region for Γ , cannot be rejected at a 95% confidence level. This is analogous to forming a 95% confidence interval for a point estimate.

²⁵This simple result comes from Corollary 4.5 of Beresteanu and Molinari (2008).

²⁶We account for the estimation error from the first stage by incorporating the bootstrapped estimates of β when forming the intervals for θ_j .