MARKETING ANALYTICS CENTRE WORKING PAPER SERIES

WORKING PAPER NUMBER:

MAC.2026.004

Title:

Who’s your Neighbor? : Measurement and Implications of Retail Agglomeration

Authors:

Hee Mok Park, Joseph Pancras

Who's your Neighbor? : Measurement and Implications of Retail Agglomeration

Hee Mok Park[1] and Joseph Pancras[2]

[1]Associate Professor of Marketing and Consumer Studies, Gordon S. Lang School of Business and Economics, University of Guelph, Canada, hmpark@uoguelph.ca

[2] Associate Professor of Marketing, School of Business, University of Connecticut, USA, jpancras@business.uconn.edu

June, 2026

Executive Summary

This study examines how grocery stores benefit from being located near non-grocery retailers. When consumers combine errands into one shopping trip, nearby businesses can influence which grocery store they choose and how far they are willing to travel. The study quantifies these benefits to inform retail location decisions.

Using shopping data from 4,749 U.S. households, the analysis compares grocery-only trips with trips involving non-grocery shopping. It accounts for household and store differences and uses seasonal variation in non-grocery shopping to help distinguish spillover benefits from the underlying attractiveness of a location.

Visiting a nearby non-grocery retailer increases the probability of choosing a grocery store by approximately 13%, on average. These benefits are concentrated within one mile and vary by neighboring retailer type. Toy stores, big-box retailers, and home improvement stores generate significant spillovers. Toy stores provide the largest benefit per shopping occasion, while big-box retailers generate the greatest overall benefit because consumers visit them more frequently. The estimated benefits translate into an expansion in effective customer reach of approximately 2.5 miles for big-box and home improvement stores and 3.9 miles for toy stores.

For retailers and shopping-center developers, the findings highlight the importance of evaluating the mix of neighboring businesses when selecting locations. Beyond population, demographics, and competition, managers should consider whether nearby retailers attract complementary shopping trips, how frequently they are visited, and their proximity to the proposed store. The study provides a framework for estimating these benefits and weighing them against additional location costs.

ABSTRACT

Retail agglomeration effects are difficult to measure due to possible reverse causality and the need to decompose agglomeration into traffic and spillover effects. In this paper we develop a new method to measure retail agglomeration in the grocery product context by exploiting consumer visits to non-grocery store formats. The seasonal variation in these non-grocery store visits contrasts sharply with the perennial and predictable nature of grocery visits. This variation enables us to establish a causal link between non-grocery visits and grocery visits and to measure the spillover effects of non-grocery visits on grocery store choice probability. We measure agglomeration effects of 13%. Our method also enables the measurement of increase in catchment area for a grocery store due to proximal location with other store formats. We find that increase in catchment area ranges from about 2.5 miles for big box and home improvement stores to 3.9 miles for toy stores, and discuss managerial implications for retailers of this method.

Keywords: Retail agglomeration; retail competition; spillover; traffic; hierarchical Bayes; causality; endogeneity

INTRODUCTION

Location is an important driver of retailer performance (Hotelling 1929, Huff 1964). Academic researchers and practitioners (e.g., Salvaneschi 1996) have alluded to proximity to dense markets with high purchasing power, availability of ample parking, and ease of access as important characteristics of attractive retail locations. In addition, researchers have suggested that positive spillover benefits might accrue from being located close to many other retailers. Such a retail agglomeration effect can arise due to savings in transportation costs for suppliers and consumers as well as labor market pooling, which enables collocated retailers to find workers (Ellison et al. 2010). Of these, the spillover benefit that accrues to co-located retailers due to consumer-side externalities is an important issue in the Marketing literature. Such consumer-side externalities can arise via two mechanisms. The first mechanism is the traffic-generating effect of retail agglomeration. This can arise because of the expectation that higher competition within the same retail format (e.g. groceries) will lead to lower prices and/or higher quality. It can also be due to lower search costs, as the presence of more retailers of the same format can also help address concerns about stock outs and long wait times; in addition multi-purpose shopping could result in greater traffic to areas with denser retail locations.[1] The second mechanism can be either positive or negative based on whether the proximal retail locations offer substitute products. If co-located retailers offer substitute products, they have to compete for customers visiting that location. Therefore, the net benefit of collocation would depend on whether the traffic generating effect exceeds the adverse consequence of increased competition. On the other hand, co-location of retailers selling non-substitute products are likely to realize the positive traffic generating effect.

In this paper, we propose an approach to measure agglomeration effect derived by grocery stores that are located close to non-grocery retailers selling non-substitute products. A common challenge in characterizing this agglomeration effect is that attractive markets (in terms of shopper characteristics) also attract multiple retailers. Researchers have attempted to circumvent this challenge by using survey data Arentze et al. (2005) or case studies involving entry of a retailer selling complementary products and services (e.g., Sen et al. (2012)). In contrast to these approaches, we exploit the seasonal variation in visits in non-grocery stores as compared to grocery stores to measure agglomeration effects across multiple grocery chains and non-grocery formats. This analysis using household-level shopping transactions enables us to measure the increase in catchment area for a grocery store as a result of being located close to a non-grocery store, a metric of key interest to retailers that is not available from the above methods.

We use the AC Nielsen Homescan panel dataset of household shopping behavior across multiple types of retailers in our empirical analysis. We infer the agglomeration effect based on the extent to which household shopping behavior differs across single- vs. multi-purpose shopping occasions. Based on model-free evidence, we document that (a) households tend to switch grocery stores on multi-purpose shopping trips, (b) they are willing to travel to grocery stores that are located farther away on such multi-purpose shopping trips, suggesting a positive externality arising from trip chaining, and (c) the average expenditures and the day of the week when the two types of grocery shopping trips occurred are very similar for single- and multi-purpose shopping occasions.

Based on this model-free evidence, we characterize the *conditional* agglomeration effect[2] as the additional benefit that a household derives from shopping at a grocery store located close to a non-grocery store *conditional* on there being a need to shop at the non-grocery store. We account for differences in the attractiveness of stores due to unobserved factors using store fixed effects and household fixed effects. A household's distance from alternative grocery stores helps explain her preference to shop at a grocery store over and above the store fixed effects. This characterizes the travel cost incurred by a customer in visiting a given store. The quantification of the travel cost helps us express the spillover effect in terms of the expansion a grocery store's captive market.

There are two econometric challenges in inferring the conditional spillover effect as the additional benefit that a household derives from shopping at a grocery store located close to a non-grocery store *conditional* on there being a need to shop at the non-grocery store. The first challenge is that correlated unobservables might have driven non-grocery demand and the need to shop at a co-located grocery store. The second challenge is establishing the direction of causality, i.e., changes in non-grocery demand led to changes in grocery store patronage. We account for these by exploiting the pronounced seasonal variation in the demand for non-grocery purchases at different types of retailers. We consider five different classes of non-grocery purchases that exhibit different seasonal patterns to aid in our identification of the agglomeration effect. In addition, we use several other variables that are likely to act as shifters of non-grocery purchases, but are unlikely to directly influence the choice of grocery stores as instruments.

We use the control function approach proposed by Petrin and Train (2010) to address the endogeneity of the decision to visit a non-grocery store. However, given that the endogenous variable, i.e., the decision to visit a non-grocery store is discrete in our context, we use the modification of the control function approach developed by Terza et al. (2008) and Wooldridge (2014) to resolve the endogeneity of our discrete explanatory variable. Also, to fully explore the richness of our individual level data, we utilize the Hierarchical Bayesian method incorporating the Metropolis-Hastings algorithm to calibrate the multinomial logit model (Allenby and Rossi 1999).

We find that grocery stores derive a statistically as well as economically significant positive spillover effect due to being located close to a non-grocery store. The spillover results in a 13% increase in purchase probability for the grocery stores, and also results in an increase in catchment area. This increase in catchment area is substantial, ranging from about 2.5 miles for big box and home improvement stores to 3.9 miles for toy stores. However, we find that the spillover effect exists only when the grocery store is located within one mile of a non-grocery store; there is no statistically significant effect beyond one mile. Among the different types of non-grocery stores, we find statistically significant effects for toy, big box, and home improvement stores. Of these grocery stores located next to toy stores experienced the highest spillover benefit *conditional* on there being a need to shop at a toy store. However, when we consider the frequency of visits to different types of non-grocery stores, we find that the resulting unconditional spillover effect is larger for big box stores.

Our analysis shows how to estimate retail agglomeration and spillover rate from scanner panel data on household store visits. Traditional retail location decisions include simple heuristics such as maximizing population enclosed within concentric rings of the location (Salvaneschi, 1996) and locating closest to the major competitor in the category, as well as more sophisticated strategies such as utilizing demographics around the location, etc. The cross-category spillovers obtained through measuring potential agglomeration effects can help retailers make location (entry/exit) decisions as well as potentially decisions on assortment choice. In summary, our paper makes a substantive contribution by quantifying the economic benefits from locating next to various types of retail stores so that managers can make optimal store location decisions. Methodologically, we propose an approach to causally infer agglomeration effect from observed retail shopping data by exploiting seasonal fluctuations in demand for non-grocery stores as instrumental variables.

The rest of the paper is organized as follows. Next we review key studies relevant to the measurement of retail agglomeration effects, provide details on the data and provide some preliminary evidence for agglomeration effects in the data. We then develop our model, discuss the identification of parameters and provide details on the estimation procedure. Finally we describe estimation results and conclusions from our study.

RELATED LITERATURE

A key challenge in empirically isolating the agglomeration effect is that shopping districts with multiple retailers are likely be located in areas that are intrinsically more attractive for businesses. Typically, intrinsic attractiveness is unobserved by the researcher. Consequently, it is often hard to tell if the higher consumer traffic results from an agglomeration effect or if it is due to the business potential in that area. In the past, researchers have tried to disentangle these alternative forces using store sales data. One approach to measuring the magnitude of agglomeration effect is based on location choices made by retailers (Datta and Sudhir 2011, Vitorino 2012). As a result, the inference of agglomeration effect is based on the assumption that firms have knowledge of the magnitude of agglomeration effect and make optimal decisions given this information. They typically account for unobserved location characteristics by invoking the notion that in equilibrium, the number of retailers in an area would inform us about its attractiveness.

Two streams of research eschew the supply side equilibrium assumptions of this approach to measuring the agglomeration effect. The first approach uses revealed preference of consumer location utility using consumer choice data to measure retail agglomeration effects.[3] Based on conjoint data, Dellaert et al. (1998) introduce a new modeling frame work that incorporates consumers' multi-purpose and multi-stop shopping behavior. Their model significantly improves the fitness of the data than the conventional logit model. Fox and McLaughlin (2007) extend the previous literature by quantifying the retail agglomeration effect. They show that consumers' amount of money spent in a store is positively correlated with the number of other stores (i.e. in different format) surrounding it. The study, however, does not take into account the possibility of correlated unobservables and reverse causality that could bias the estimates. To overcome such issues, Sen et al. (2012) use a difference-in-difference approach to measure the agglomeration effect of the gas station on grocery purchases of those visiting the gas station and those that do not. After controlling for category purchases they find that the gas station results in an increase of about 8% in grocery purchases. The major parallel between our study and Sen et al. (2012) is that both studies focus on measuring the agglomeration effect using trip chaining panel data of consumers where the focal store format is grocery stores. But there are several important differences. First, Sen et al. (2012) measure agglomeration due to store entry in the form of a complementary gas station format by an existing (grocery) chain. This will arguably lead to market expansion effects for the specific grocery store chain, whereby consumers switch from their existing gas station to the new gas station. This switch in gas station patronage should lead to increase in grocery purchases due to spillover effects, and measuring this effect was the focus of Sen et al (2012). By contrast our approach to measuring agglomeration effect does not rely on store entries which tend to be strategic and somewhat lower frequency events in a mature market. Second, our application enables us to draw on agglomeration across different types of non-grocery outlets such as electronics and toys, including categories commonly used as anchor stores in malls and therefore a source of traffic from destination shopping or trip chaining. Third, Sen et al. (2012) measure the agglomeration as the change in expenditure rather than store choice. By studying changes store choice due to trip chaining, we measure the potential increase in catchment area due to agglomeration, which can be an important metric for retail managers to track.

The second approach to measuring agglomeration effect relies on stated (as opposed to revealed) preference to shop at co-located retail locations during multipurpose shopping occasions. Leszczyc et al. (2004) and Arentze et al.(2005) conducted empirical studies on how the agglomeration of stores increases consumer preference for that location or store. Their approach is based on estimation of a nested logit model using survey data in which they asked what grocery store (e.g., grocery store near a mall vs. grocery standing alone) respondents chose for multipurpose shopping versus single-purpose shopping. While these studies provide insightful approaches to exploring the existence of agglomeration effect, the data based on phone survey could face limitations such as the use of the last purchase occasion only and the use of limited number of categories (groceries, clothing and shoes and others) to characterize multi-purpose shopping. We view our measurement of agglomeration effect as addressing several of these limitations in that we use panel data of actual transactions and an explicit causal framework to measure the agglomeration effect; thus our study complements and builds on the insights of Leszczyc et al. (2004) and Arentze et al. (2005).

DATA

We use the ACNielsen Homescan data to understand the role of retail agglomeration in a household's choice of grocery shopping destination. The data includes visits and purchase histories of 4,749 households living in 34 states in grocery and non-grocery stores between January 1998 and December 2002. In addition to visit and purchase information, the data also include information on the ZIP code where each household lives as well as the ZIP code of the grocery stores that they visit. We use this information to calculate the distance traveled for each grocery shopping trip. Specifically, we use the geocode (i.e., longitude and latitude) for each ZIP code and use the geocode to calculate the Euclidean distance between two ZIP codes. However, our data do not have the exact location information for non-grocery stores visited by each household. Nevertheless, we know the name of the store that was visited (i.e., brand name such as Best Buy). We imputed the location information for the non-grocery stores by identifying the nearest store with that brand using a data from Trade Dimensions TDLINX 2002 data base.

We present the descriptive statistics of variables of interest regarding grocery shopping by households in Table 1. On average, households visited 0.25 stores per day, which translates to about one to two stores during a week. In a typical store visit, households spent $39.16 and traveled an average of 3.5 miles for each store visit. We further analyze these variables in next section to see if there is any model-free evidence supporting retail agglomeration effects.

"Insert Table 1 about here"

Model-free evidence of agglomeration effect

Agglomeration effect from co-location of retail outlets can arise because of two reasons: (a) mere presence of multiple retailers can render a location attractive (or plausibly unattractive) for shopping and (b) in instances when there is a need to shop at multiple retailers, trip-chaining can render collocated retailers attractive for shopping. Of these, (a) is difficult to empirically isolate because locations that are patronized by consumers with desirable characteristics (e.g., purchasing power) tend to attract multiple retailers. Therefore, we restrict our analysis to empirically isolating (b). Subsequently, we discuss the role of (a) in the retail location choice decision.

In our empirical analysis, we investigate if grocery stores located closer to non-grocery stores derive such a positive spillover. If they do, in instances where households experience a need to shop at a certain non-grocery store, they should prefer to patronize a grocery stores located close to it. In some instances, this might lead to switching away for a household's favorite store in favor of a store that is located closer to the non-grocery store that is visited. In addition, Thomadsen (2005) and Huff (1964) have documented that travel cost is an important consideration when customers decide on a retail location. Consequently, all else equal, households would prefer a more proximate retail location. If multipurpose shopping results in positive spillover benefits, a grocery store located close to non-grocery stores should be able to draw customers from a wider geographic area. As a corollary, customers should be willing to travel to a grocery store that is farther out on multipurpose shopping occasions than they do on solitary grocery shopping trips.

In order to investigate whether households switch grocery stores during multipurpose shopping occasions, we considered the average distance traveled for a grocery store in both cases. We find that an average household traveled 3.19 miles to visit a grocery store when it was a single purpose shopping occasion and 3.62 miles for multipurpose shopping. This suggests that households are willing to travel to a different grocery store on multipurpose shopping occasions and also patronize a grocery store that is farther away on such trips.

However, a potential concern with this analysis is that households that indulge in multipurpose shopping might be inherently different in terms of distance traveled to grocery stores. If this is true, longer distance traveled for grocery shopping on multipurpose shopping occasions might be a household characteristic rather than a result of multipurpose shopping per se. We control for this possibility by regressing the distance traveled to a grocery store on a dummy variable that indicated whether the trip was considered multipurpose shopping and household fixed effects. The household fixed effects account for cross-sectional differences across individuals in terms of the average distance traveled for a typical grocery-shopping trip. This enables us to identify within-household changes in distance traveled to visit a grocery store on single versus multipurpose shopping occasions. The results from this analysis suggest that multipurpose shopping trips had a positive and significant effect on the distance traveled to a grocery store. Specifically, households have tendency to visit a grocery store located 0.407 miles (13%) farther when they were engaged in multipurpose shopping than single-purpose shopping.

We further investigate if the nature of single-purpose and multi-purpose grocery shopping visits differ on other meaningful dimensions such as the average expenditure and the day of the week when the trip was made. We find that there is no statistically significant difference in the average expenditure on the two types of grocery shopping trips: $33.61 on single-purpose trips vs. $36.31 on multi-purpose visits. Similarly, 34% of single-purpose and 41% of multi-purpose grocery visits occurred during weekends. Once again, this difference is not statistically significant. On the other hand, we find that the grocery store visited during multi-purpose shopping trips are located closer to the other non-grocery stores (average distance of 5.32 miles) than those patronized during single-purpose shopping trips(average distance of 6.62 miles). This difference is statistically significant at the 0.01 level.

Together, there are three key highlights from these data patterns. First, households are willing to switch patronage of grocery stores on multi-purpose shopping trips. The grocery stores patronized during multi-purpose shopping trips are located close to non-grocery stores where households shop. Second, households are willing to travel to grocery stores that are located farther away on such multi-purpose shopping trips, suggesting a positive externality arising from trip chaining. Third, the average expenditures and the day of the week when the two types of grocery shopping trips occurred are very similar for single- and multi-purpose shopping occasions. Together, these highlights suggest that although households tend to switch grocery stores on multi-purpose shopping trips, single- and multi-purpose grocery shopping trips do not differ in any meaningful way (e.g., fill-in vs. regular trips (Bell and Lattin 1998)). Stated differently, these patterns suggest that the only difference between the two types of trips is the choice of grocery store conditional on there being a need for such a visit.   Below, we develop our empirical model that is consistent with these data patterns.

EMPIRICAL SPECIFICATION

Model and estimation

Given that the model-free evidence suggests that while households switch grocery stores on multi-purpose shopping occasions, but do not alter the nature of grocery shopping, we consider the choice of grocery store conditional on there being a need to make such a trip. For the sake of estimation, we restricted that a household only considers grocery stores within a reasonable distance from his/her home which we set as fifty miles. We also excluded those grocery stores that were visited less than one time per year. This resulted in the average size of the consideration set to be four grocery stores per household. We specify the utility that household *i* derives from visiting grocery store *j* at time *t* (day) as follows:

$$U_{ijt}^{GR} = \alpha_{0i} + \alpha_{1i} D_{ij} + \sum_{k} \alpha_{2ki}\ I_{ijt}^{k} + sf_j + \eta_{ijt}. \qquad (1)$$

The utility specification in Equation 1 includes two important drivers of a household's decision to visit a particular grocery store: the grocery store's proximity to the consumer (D_{ij}) and the household's need to shop at one or more non-grocery stores that happen to be in the vicinity of grocery store *j* (I_{ijt}^{k}). The term D_{ij} is the Euclidean distance between the household *i* and grocery store *j*. α_{1i} captures the effect of distance between grocery store *j* and household *i* on utility (and hence choice probability for store *j*). If households perceive a travel cost, they would prefer more proximate grocery stores. Consequently, we would expect α_{1i} to be negative. The term I_{ijt}^{k} is a binary variable that indicates whether household *i* visited the non-grocery store of type *k*, *k* = 1, 2, …, *K*, located close to grocery store *j* at time *t*. α_{2ik} is the main parameter of interest and represents the additional benefit that grocery store *j* derives as a result of household *i* visiting the non-grocery store of type *k* in its vicinity. The coefficients α_{1i} and α_{2ik} are estimated at an individual level using Hierarchical Bayesian method (Allenby and Rossi 1998) which allows us to explore heterogeneity of the travel cost and agglomeration effect across households. In our application, we consider *K* = 5 types of non-grocery stores: (i) toy stores such as Toys R Us, (ii) electronic stores such as Best Buy, (iii) big box retailers such as Walmart and Target, (iv) home improvement stores such as Home Depot and Lowes, and (v) office supply stores such as Staples and Office Depot.

Given that our focus is on measuring spillover effects between grocery stores and retailers that are not substitutes for grocery shopping, we excluded supercenter versions of big box retailers.[4]

The term sf_j is the store fixed effect for store j, which captures the time invariant factors that drive the attractiveness of the store. The term α_{0i} is the household fixed effect for household i. Recall that we constructed household's consideration set by stores that meet two criteria; whether the distance from the household's home is less than fifty miles and whether the household visited the store more than once a year. We included household fixed effects α_{0i} that capture average intrinsic utility (time invariant) that the person has toward other grocery stores compared to the baseline store.[5] These store fixed effects sf_j capture the average popularity the store has across households after accounting for the differential impact from different baseline stores through household fixed effects α_{0i}.

The term η_{ijt} is the error term. We decompose the error term into two components such that $\eta_{ijt} = \xi_{ijt} + \varepsilon_{ijt}$. As we discuss subsequently, the first component, ξ_{ijt} represents time variant factors unobserved by the researcher that could impact household i's grocery and non-grocery store visit decisions. As a result, the ξ_{ijt} component of η_{ijt} could be correlated with I_{ijt}^{k}, which in turn would lead to biased parameter estimates.

In order to resolve this endogeneity issue, we follow the approach proposed by Terza et al. (2008) and Wooldridge (2014), which is a modification of the control function approach proposed by Petrin and Train (2009) to accommodate discrete endogenous variables. The basic procedure is similar to the control function approach, wherein the residual from the first stage model is used as an extra explanatory variable in the main model. The first stage model is set up as follows.

$$U_{ijt}^{k} = \beta_i^{k\prime} Z_{ijt}^{k} + \tau_{ijt}^{k} \qquad (2)$$

U_{ijt}^{k} is a utility that a household *i* has toward a non-grocery store of type *k* located within one mile distance to the grocery store *j*. Note that when $U_{ijt}^{k} > 0$, I_{ijt}^{k} in Equation (1) is 1 and is 0 otherwise. Z_{ijt}^{k} is a vector of explanatory variables that includes the instruments that affect the household's demand for the non-grocery store independent from his/her demand for the grocery store *j*.[6] Assuming τ_{ijt}^{k} follows standard normal distribution, Equation (2) is a probit regression. We then obtain the residual estimates as follows.

$$\widehat{\tau_{ijt}^{k}} = \left(I_{ijt}^{k}\right) - \Phi\left(\widehat{\beta_i^{k}}' Z_{ijt}^{k}\right), \qquad (3)$$

Note that we include single-purpose visits to these non-grocery stores as well as those that are combined with grocery store visits in the first stage model. $\widehat{\beta_i^{k}}$ is the vector of parameter estimates for the first stage probit regression. Φ is standard normal cdf. As shown below, we plug in $\widehat{\tau_{ijt}^{k}}$, k= 1, 2, …, 5, as extra explanatory variables to capture the correlation between ξ_{ijt} and I_{ijt}^{k}. Assuming ε_{ijt} follows i.i.d. type I extreme value distribution, the estimation boils down to calibrating the conditional logit model as follows.

$$U_{ijt}^{GR} = \alpha_{0i} + \alpha_{1i} D_{ij} + \sum_{k} \alpha_{2ki} I_{ijt}^{k} + sf_j + \sum_{k} \alpha_{3i} \widehat{\tau_{ijt}^{k}} + \varepsilon_{ijt} \qquad (4)$$

The observed choice y_{it} of household *i* is

$$y_{it} = j \;\; if \;\; U_{ijt}^{GR} > U_{ij\prime t}^{GR} \;\; for\ all\ j' \qquad (5)$$

The probability that household *i* chooses grocery store *j* is

$$\mathrm{P}(y_{it} = j |\, \tilde{X}_{i1t}, \tilde{X}_{i2t}, \cdots \tilde{X}_{iJt}) = \frac{\exp(\alpha_i' \tilde{X}_{ijt})}{\sum_{j'} \exp(\alpha_{i_i}' \tilde{X}_{ij't})} \qquad (6)$$

where $\tilde{X}_{ijt} = (1, D_{ij}, \sum_k I\ (U_{ijt}^k > 0), sf_j, \widehat{\tau_{ijt}^1}, \ldots, \widehat{\tau_{ijt}^5})$

For both probit and multinomial logit regression we used maximum likelihood estimation. We used two stage estimation where we corrected the standard errors using the method proposed by Terza (2012a). To fully explore the richness of our individual level data, we used Hierarchical Bayesian method (Allenby and Rossi 1998) which allows us to estimate parameters at an individual level. We applied Gibbs sampling method for estimating probit model and incorporated the Metropolis-Hastings algorithm to calibrate the multinomial logit model. We present the detailed procedure and prior information for Hierarchical Bayesian approach in the Appendix A.

Identification

There are three main challenges in empirically identifying the agglomeration effect, α_{2ik} based on observational data. The first challenge is that the preference for a grocery store might be driven by the unobserved intrinsic attractiveness of the location. Consequently, locations that households find attractive might also draw a wide variety of retailers. Thus grocery stores located in close proximity to many non-grocery stores might be preferred because of unobserved location characteristics. Since such attractiveness is not observed by researchers, it is often challenging to determine if higher consumer traffic for both stores results from the spillover effect or from business potential of the area (Orhun 2013). We utilize the disaggregated and panel nature of our data to control for location-specific unobservable factors. The panel structure of the data allows us to apply store fixed effects, sf_j, to capture the time invariant unobserved drivers of the utility of shopping at grocery store j'. Consequently, α_{2ik} will pick up how visits to a given store change based on visits made to a nearby non-grocery store of type k, i.e., I_{ijt}^{k}.

The second challenge is that we need to have sufficient variation in the extent to which grocery stores are located in close proximity to non-grocery stores in order to identify the agglomeration effect. More importantly, we also need variation in the extent to which households indulge in multi-purpose grocery shopping trips involving different combinations of non-grocery store types. In essence, if all the grocery store visits are either single-purpose trips or involve only a multi-purpose shopping trips with only a few types of non-grocery stores located nearby, we cannot identify the agglomeration effect. Therefore, we explore whether the grocery stores in our data have sufficient variation in terms of being located close to different types of non-grocery stores. Furthermore, since we propose to identify different spillover effects by store format, we need variation in the configuration of grocery stores by format.

In Tables C1 and C2 in the appendix, we present the proportion of grocery stores and store visits with different combinations of non-grocery stores within one mile. These results suggest that roughly 32% of grocery stores are in locations that do not have any non-grocery store in their vicinity, while the remainder are collocated with non-grocery stores. Similarly, 29% of visits are to stores that do not have a non-grocery store in their neighborhood. These patterns suggest that we have sufficient variation to identify the spillover effect of visiting a non-grocery store in the neighborhood of a grocery store. Let us now consider how we can identify the spillover effect by the format of non-grocery retailer. The results in Tables C1 and C2 in the appendix suggest that we have a sizable number of stores and visits that differ in terms of the combinations of non-grocery stores in their neighborhood. This variation, in addition to our assumption regarding the additive nature of the spillover effect from different non-grocery store will help in identify the spillover effect by format.

The third issue in identifying the agglomeration effect is that visits to non-grocery stores are endogenous decisions made by the household. In particular, time varying factors not observed by researchers could drive a consumer to visit both stores during the same time. For example, special occasions such as a child's birthday might induce visits to a toy store and a specialty grocery store in its vicinity. In this case, the household patronized the specialty grocery store not because it was located near a toy store, but rather, a common unobservable, child's birthday drove both visits. In addition, recall that we investigate the spillover effect that accrues to grocery stores due to being located close to a non-grocery store. We attempt to infer this effect by considering how visits to a nearby non-grocery store shift a household's utility from visiting a grocery store. One can argue that even if such multipurpose shopping trips involving visits to nearby grocery and non-grocery stores exists, the direction of the spillover effect is not clear. For instance, an exogenous need to visit a specialty grocery store that the household does not regularly patronize might have induced the buyer to utilize the opportunity and make non-grocery purchases from stores nearby. Under such a scenario, grocery store choice led to non-grocery purchases rather than the other way around.

In order to address these issues associated with correlated unobservables and to establish the direction of causality (i.e., non-grocery visits driving grocery store choice), we use exogenous factors (i.e., instruments) that can serve as valid exclusion restrictions. Specifically, these instruments should shift visits to non-grocery stores $(i.e., I_{ijt}^{k})$ but should be independent of grocery demand. In our application, we identify four classes of instruments. The first class of instruments relies on the idea that the different types of non-grocery stores exhibit their own unique patterns of seasonality. In Figure 1, we present the average number of visits by the month of the year for grocery stores as well as the five types of non-grocery stores.[7] These figures suggest that there is limited seasonality in grocery store visits. The highest deviation from the average number of visits was in March which was within 5% of those in other months. By contrast non-grocery stores exhibit a large degree of seasonality with different types of non-grocery stores experiencing their peak traffic during different times of the year. For example, toy, electronic, and big box stores experience significant increase in store visits during the holiday gift shopping season (November and December). In fact, households were 20% more likely to visit these stores during November and 85% more during December than in other periods. On the other hand, home improvement stores see an increase in visits during spring (May), while office product stores see 21% higher traffic during back-to-school periods (August, September, December, and January). Thus, we argue that while seasonality is a driver of non-grocery visits, it does not have an effect on demand for grocery products and can thus be a valid instrument. Furthermore, since alternative forms of non-grocery stores exhibit different patterns of seasonality, we are likely to have sufficient variation to allow the instruments to have high explanatory power. We name this variable as $Seasonal\_Demand_{it}$.

"Insert Figure 1 about here"

The second instrument is based on the idea that there are cross-sectional differences across households in terms of their intrinsic propensity to visit non-grocery stores of each type. We operationalize this via the estimated household-specific intercept for each non-grocery store category. Once again, the rationale for this instrument is that higher demand for a particular non-grocery is unlikely to impact the need to visit grocery stores.

The third instrument that we utilize is the household-specific seasonality in demand for each type of non-grocery store over and above the common seasonality across households in the first instrument. To be specific, we use the average number of visits a household made to a non-grocery category during a specific week in the previous years. The instrument captures a household's idiosyncratic demand for a type of non-grocery store during periods over and above what is picked up by the average (across households) seasonal fluctuation in non-grocery demand. The rationale behind the exclusion restriction is that the preference for a particular grocery store, however, does not necessarily change during that period. Note that the dependent variable in the first stage and the instrument include both solitary trips to non-grocery stores as well as those combined with grocery store visits. The inclusion of solitary trips helps in capturing seasonal demand for non-grocery purchases within a household even when there is no need for a concomitant grocery purchase. We name this variable as $Non\_Seasonal\_Demand_{it}$.

The last instrument is the number of times a household visited a non-grocery store during the previous three weeks. The variable captures the household's state dependence toward non-grocery demand, although it is unlikely to be directly related to the choice of grocery store. For instance, a consumer who had a new born baby will start visiting a baby product store frequently. We name this variable as $Pre\_Three\_Wks_{it}$.

In summary, store fixed effects and the instrumental variables enable us to pin down the causal effect of demand for non-grocery stores on the choice of a grocery store. Incorporating the instruments we can specify the first stage probit regressions model as follows.

$$U_{ijt}^{k} = \beta_{0i}^{k} + \beta_{1i}^{k} D_{ij}^{k} + \beta_{2i}^{k} Seasonal\_Demand_{it} + \beta_{3i}^{k} Non\_Seasonal\_Demand_{it} + \beta_{4i}^{k} Pre\_Three\_Wks_{it} + \tau_{ijt}^{k} \qquad (7)$$

where $k \in \{\text{Toy, Electronic, Big Box, Home improvement, Office product}\}$. D_{ij}^{k} is the distance between household *i* and the non-grocery store of type *k* located near grocery store *j*.

A common concern with using instrumental variables is that they may not be strongly correlated with the endogenous variable, which in our context, is the visits to non-grocery stores. Such weak instruments can lead to biased parameter estimates (Rossi 2014). In order to verify that we do not have a weak instruments problem, we estimate the first stage model with a household's visits to non-grocery store as the dependent variable and the exogenous variables and the instruments as independent variables. In essence, a comparison of the predictive power of instruments over and above that of the exogenous variables is an indicator of the strength of the instruments.

We present the results from the first stage analysis in Table 2. Given that our endogenous variable is discrete, we estimate binary probit models to characterize visits to non-grocery stores of each type. The results show that coefficients for all variables including excluded variables (i.e., instrumental variables) were statistically significant with expected signs. Coefficients for $Seasonal\_Demand_{it}$ was positively significant meaning that consumer had significantly higher demand for non-grocery products during their popular seasons. The coefficients for

$Non\_Seasonal\_Demand_{it}$ was also positive and significant implying that consumer-period specific demand exists when they have higher demand non-grocery products than other periods (e.g., consumer's birthday). Finally, positive coefficients for $Pre\_Three\_Wks_{it}$ indicates that positive state dependence exists for consumers' demand for non-grocery products.

"Insert Table 2 about here"

In order to assess the explanatory power of the instrumental variables separate from the other exogenous variables, we report Mcfadden's Pseudo R-squared for each non-grocery category with excluded variables versus without them. Specifically, in the second row of Table 3 below ("with excluded variables model") we add the four instruments to the model in the first row ("without excluded variables model") which has only homogeneous intercept and distance variable. The result in Table 3 shows that our instruments significantly improve the explanatory power of the model.

"Insert Table 3 about here"

Results

In this section, we discuss the estimation results which is based on a sample of 1415 households living in the states of New York, Indiana, and Illinois.[8] Recall that we characterize the spillover effect as the increase in utility that accrues to a grocery store by virtue of a household shopping at a non-grocery store in its vicinity. Also recall that the variation in seasonality across store types as well as the diversity in retail configuration with different store types enable us to infer separate spillover effects for each type of non-grocery store. Nevertheless, we begin by discussing the results from a model with a common agglomeration coefficient for all non-grocery categories, i.e., $\alpha_{2ik} = \alpha_{2i}, k = 1,2, \dots ,5$. Subsequently, we discuss the results with heterogeneous spillover effects generated by visits to different types of retailers.

In order to characterize the spillover effect, we need to define what constitutes a proximate non-grocery store. In Table 4, we present the results of the population estimate with proximity defined as a distance of less than one mile. The results reveal that the coefficient for D_{ij} is negative and statistically significant, suggesting that households incur a travel cost when they make trips to grocery stores. The estimate implies that, all else equal, an average household is 5.01% less likely to visit a grocery store that is located one mile farther away compared to another equivalent grocery store. The main parameter of interest, the coefficient of I_{ijt}^{k} , is positive and statistically significant. This suggests that households derive a positive benefit from visiting a grocery store that is located within one mile of a non-grocery store in instances when there is a need to visit the said non-grocery store.

In order to put this spillover effect in perspective, we characterize the benefits that might accrue from this externality. The first benefit that accrues from the spillover effect is that a visit to a non-grocery store would increase the likelihood that the household switches to a grocery store that is located within one mile. Note that since we have included store fixed effects, this benefit is over and above any other cross-sectional benefit that might accrue due to unobserved location attractiveness. The second, albeit related benefit, is tied to the travel cost, which imposes a disutility for grocery stores located farther away from a household. The positive spillover effect would dampen this negative effect and increase the catchment area of a grocery store when there is need for non-grocery shopping. Consequently, a grocery store located close to a non-grocery store has the potential to draw customers from a larger area when they visit a non-grocery store in its vicinity.

"Insert Table 4 about here"

"Insert Table 5 about here"

In Table 5, we present the agglomeration effects in terms of the increase in choice probability that a grocery store would realize when a household visits a non-grocery store located within one mile. In addition to the change in choice probability for an average store, we also present this effect as a function of the popularity of the grocery store, as measured by the estimated store fixed effect. The first column (No Non-Grocery Visit) represents the baseline choice probability when there is no visit to a proximate non-grocery store. Similarly, the second column (Non-Grocery Visit<1mile) represents the choice probability when there is a visit to a non-grocery store in its vicinity. The last column shows the corresponding percentage increase in choice probability. These results suggest that spillover effect increases the choice probability by approximately 2-3% points, which corresponds to 12-15% change in choice probabilities.

We can also infer how this agglomeration effect increases the catchment area conditional on a household visiting a proximal non-grocery store. From Table 4, the coefficient of D_{ij} indicates that one mile increase in distance decreases consumer's utility by 0.275. Given that the coefficient of I_{ijt}^{k} is 0.165, we can infer that the catchment area increases by 0.6 miles (0.165/0.275=0.6) as a result of agglomeration. Specifically, a consumer is willing to travel 0.6 miles farther to visit a grocery store that is near a non-grocery store in demand than other grocery stores.

It is conceivable that the travel cost that households incur is likely to be a function of traffic in the neighborhood. Therefore, households who live in more urbanized areas with higher population density could be more willing to visit stores that are co-located each other. In order to understand if the magnitude of agglomeration effect is bigger for households residing in more urbanized areas than those in less urbanized areas, we considered the differences in the individual level coefficients of I_{ijt}^{k} across the two types of areas. We defined more urbanized areas as cities with population density higher than 7,000 people per square mile (median in our sample) and rest of the cities as less urbanized area. Individual coefficients of I_{ijt}^{k} were regressed on a dummy variable that took on a value of 1 for more urbanized area and 0 for other areas. The result from the analysis suggest that households living in more urbanized areas did not necessarily show higher magnitude of agglomeration effect as the coefficient of the dummy variable was statistically insignificant (estimate = -0.009, standard error = 0.026).

Sensitivity to alternative definitions of proximity

Recall that the above analysis is based on a pre-determined definition of proximity. In what follows, we relax this definition with two objectives: (a) to check the sensitivity of the results to alternative definitions of proximity and (b) to determine the boundary conditions when the spillover effect might cease to exist. To this end, we estimated Equation 1 using three alternative definitions of proximity between the grocery and non-grocery stores (while defining the dummy variable I_{ijt}^{k}): within 1 mile, within 2 miles, and within 3 miles. We present the results from this analysis in Table 6. These results suggest that the estimated agglomeration effect decreases as the definition of proximity between the grocery and non-grocery stores increases. Furthermore, the agglomeration effect becomes statistically insignificant when the definition of proximity is based on a distance of 2 miles or more. Therefore, we conjecture that the positive spillover effect realized by a grocery store by virtue of being located close to a non-grocery is bound to be local.

"Insert Table 6 about here"

In order to determine the boundary conditions for the spillover effect, we estimated a more general version of Equation 1 with multiple indicators for visits to non-grocery stores located in three distance bands (i.e. 0-1 mile and 1-2 miles) as follows:

$$U_{ijt} = \alpha 1_j + \alpha 2_i + \alpha 3_i Dist_{ij} + \alpha 4_i I(0 < Visit_{Dist\,jk} < 1)$$
$$+\alpha 5_i I(1 < Visit_{Dist\,jk} < 2) + \varepsilon_{ijt}. \quad (9)$$

Note that we restrict the spillover effect to be zero when a grocery store is located more than two miles away from a non-grocery store that is visited. We present the results from this model in Table 13. These results indicate that the agglomeration effect is positive and significant only when the non-grocery store is located within one mile of a grocery store. This reinforces our earlier conjecture that the agglomeration effect is predominantly a local phenomenon. Therefore, we restrict subsequent discussions to estimates from models where we define proximity as being located within one mile of another store.

"Insert Table 7 about here"

Heterogeneous agglomeration effect by retailer type

We present the results from the model that allows for the spillover effect to vary across these five types of retailers in Table 8. These results suggest that visits to all five types of non-grocery stores had a positive impact on the grocery stores in the vicinity, although the effect is statistically significant only toy, big box, and home improvement stores. Based on these estimates, we report the spillover effect in terms of the increase in choice probability in Table 9. These results suggest that largest positive spillover effect was realized when the household visited a toy store, which translates to a 38.46% increase in choice probability a nearby grocery store. In terms of catchment area, an average household is willing to travel up to 2.03 miles farther for a grocery store that is located near a toy store if they have a need to shop at that toy store. Spillover effect from home improvement store and big-box store were 26.11% and 25.19% while their effects in terms of catchment areas are 1.39 miles and 1.31 miles respectively. Note that these increases in catchment radius are larger than the 0.6 miles that we inferred based on common spillover effects across the different types of retailers.

"Insert Table 8 about here"

"Insert Table 9 about here"

Discussion

The results from our analysis imply that grocery stores derive a statistically significant, albeit localized spillover effect from being located close to non-grocery stores. At a first glance, this result seems to suggest that grocery retailers probably stand to gain by being located close to toy, big box, and home improvement stores. In this section, we discuss some important considerations while relating these results to the retail location choice decision.

The first consideration is that the spillover effect discussed thus far represents the additional benefit that a grocery store derives conditional on a household visiting a particular non-grocery in its vicinity. In order to arrive at a managerially meaningful metric, we need to quantify the overall unconditional benefit by accounting for the extent to which the households visit each type of non-grocery store. To this end, we consider the average number of visits to each type of non-grocery stores by the households in our data. The second column (Spillover Rate) is the increase in probability in terms of percentage point calculated from Table 9 (e.g., Toy Spillover Rate 9.25%=33.30%-24.05%). By multiplying the conditional spillover rate and the traffic per store, we obtain the number of households that a grocery store gained from locating next to the non-grocery store. We then obtain the unconditional agglomeration effect in percentage terms by dividing this number by average number of households in our sample that visited an average grocery store during the sample period (presented in the third column of Table 10). The results in Table 10 reveal that although toy stores provided the highest conditional increase in choice probability for a grocery store, they do not yield the highest unconditional benefit. On the other hand, big-box and home improvement stores, by virtue of their higher store traffic generate the greater overall agglomeration effects.

"Insert Table 10 about here"

The second consideration, as discussed earlier, is that our empirical strategy for identification is based on the extent to which the choice probability of a grocery store changes when a household visits a non-grocery store in its vicinity. Given our desire to obtain clean causal effects, we included store fixed effects to account for unobserved intrinsic attractiveness of individual grocery stores. It is conceivable that some of this intrinsic attractiveness of grocery stores is related to the presence of other retailers in its vicinity, even if a household does not visit any of these non-grocery stores. In the absence of trip chaining, collocation of grocery and non-grocery stores can have a positive or negative effect on the intrinsic attractiveness of the former. The case for the positive effect can be made based on the attractive ambience created by the presence of multiple retailers. On the other hand, increased traffic might render such a location less attractive. If collocation creates positive (negative) externalities for grocery retailers even in the absence of trip chaining, our estimate of the spillover effect should be viewed as the lower bound (upper bound) of the overall agglomeration effect.

In order to understand if the intrinsic attractiveness of stores differs across collocated versus stand-alone grocery stores, we considered the differences in the estimated store fixed effects across the two types of grocery stores by regressing the store fixed effects on a dummy variable that took on a value of 1 for stand-alone grocery stores and 0 for stores collocated with non-grocery stores. The results from this analysis suggest that the coefficient of the stand-alone dummy was statistically insignificant (estimate = 0.018, standard error = 0.047). Therefore, the spillover effect that we estimated based on instances when households indulged in multi-purpose shopping is a reasonable proxy for the total agglomeration effect. However, a potential caveat is that the store fixed effects probably comprise of three components: (a) attractiveness of a grocery store as a result of being located close to non-grocery stores even in the absence of multi-purpose shopping, (b) intrinsic propensity for grocery shopping among residents in the neighborhood, and (c) unobserved store characteristics such as cleanliness. Of these, agglomeration effect is related to (a). Therefore, differences in these fixed effects, or lack thereof, are likely to be imperfect indicators of the cross-sectional component of the agglomeration effect.

The third consideration is the cost of collocation. Our analysis provides a starting point in terms of quantifying the benefits of collocation. However, the final location choice decision needs to be made based on whether the benefits from colocation are commensurate with any additional costs. Given that we do not have information on the cost of colocation, we cannot comment on the merits of collocating grocery stores with non-grocery stores. Nevertheless, our approach can help managers who have information regarding the cost of colocation make calibrated location decisions.

CONCLUSION

Agglomeration effect is an important consideration when managers make retail location choice decisions. Typically, measuring agglomeration effect can be tricky because colocation of retail outlets can also be driven by unobserved demand characteristics in the market. This renders it difficult to infer the effect of colocation on demand. We present an approach to quantify agglomeration effect between retailers that offer broadly non-substituting products – grocery stores and non-grocery retailers. We exploit the extent to which households patronize different grocery stores on single- vs. multi-purpose shopping trips to infer the agglomeration effect.

Our empirical analysis helps in quantifying the agglomeration effect both in terms of the extent to which collocated grocery retailers can generate more traffic on multi-purpose shopping occasions and increase the radius of the captive market. However, this benefit from colocation would also depend on the type of non-grocery retailers. Specifically, our analysis suggests that there is benefit to collocating with some types of retailers such as big box retailers, toy stores, and home improvement stores. Furthermore, any benefit that grocery stores derive from collocating next to these non-grocery stores appears to be fairly local.

While these results provide some valuable insights regarding the retail location choice decision, there are some important limitations and caveats. First, as discussed above, we capture the agglomeration effect that arises from trip-chaining. Nevertheless, grocery stores can derive positive (or negative) benefits from being located close to non-grocery retailers even on single-purpose shopping occasions. While exploratory analyses suggest that the role of such cross-sectional agglomeration effect can be limited, pinning this effect is a worthwhile avenue for future research. Second, with the growing popularity of online retailing and big box retailers such as Walmart and Target stocking grocery products, the nature of agglomeration effect might have changed over time. It will be interesting to augment current analyses with those based on more recent data to understand if the nature of agglomeration effect has indeed changed over time.

REFERENCES

Allenby, Greg M., and Peter E. Rossi. “Marketing models of consumer heterogeneity.” Journal of Econometrics 89 (1999): 57-78.

Arentze, Theo A., Harmen Oppewal, and Harry JP Timmermans. "A multipurpose shopping trip model to assess retail agglomeration effects." Journal of Marketing Research 42, no. 1 (2005): 109-115.

Bell, David R., and James M. Lattin. “Shopping Behavior and Consumer Preference for Store Price Format: Why “Large Basket” Shoppers Prefer EDLP.” Marketing Science 17, no. 1 (1998): 66-88

Datta, Sumon, and K. Sudhir. "The Agglomeration-Differentiation Tradeoff in Retail Location Choice.” Yale University, mimeo (2011)

Dellaert, Benedict G.C.,Theo A. Arentze, Michel Bierlaire, Aloys W. J. Borgers and Harry J. P. Timmermans. “Investigating Consumers' Tendency to Combine Multiple Shopping Purposes and Destinations.” Journal of Marketing Research 35, no.2 (1998):177-188.

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. "What causes industry agglomeration? Evidence from coagglomeration patterns." *American Economic Review* 100, no. 3 (2010): 1195-1213.

Fox, Edward J., and Amanda McLaughlin. "The impact of retail location on retailer revenues: An Empirical investigation." Southern Methodist University, mimeo (2007).

Hotelling, Harold. “Stability in Competition.” The Economic Journal 39 (1929): 41-57.

Huff, David L. “Defining and Estimating a Trading Area.” Journal of Marketing 28, no. 3 (1964): 34-38.

Leszczyc, Peter TL Popkowski, Ashish Sinha, and Anna Sahgal. "The effect of multi-purpose shopping on pricing and location strategy for grocery stores." Journal of Retailing 80, no. 2 (2004): 85-99.

Orhun, A. Yeşim. "Spatial differentiation in the supermarket industry: The role of common information." Quantitative Marketing and Economics 11, no. 1 (2013): 3-37.

Pancras, Joseph, S. Sriram, and V. Kumar. "Empirical investigation of retail expansion and cannibalization in a dynamic environment." Management Science 58, no. 11 (2012): 2001-2018.

Petrin, Amil, and Kenneth Train. “A Control Function Approach to Endogeneity in Consumer Choice Models.” Journal of Marketing Research 47, no. 1 (2010): 3-13.

Progressive Grocer (2016), http://www.progressivegrocer.com/departments/frozen/cpg-manufacturers-show-price-declines-16-percent-2016?cc=17, accessed April 4, 2017.

Raphael, Thomadsen "The Effect of Ownership Structure on Prices in Geographically Differentiated Industries." The RAND Journal of Economics, 36, no. 4 (2005): 908-929.

Roberts, Gareth O. and Jeffrey S. Rosenthal. "Optimal scaling for various Metropolis-Hastings algorithms." Statistical Science 16, no. 4 (2001): 351-367.

Rossi, Peter "Even the Rich Can Make Themselves Poor: a critical examination of the use of IV methods in marketing." *Marketing Science, vol. 33, no. 5* (2014): *655-672.*

Salvaneschi, Luigi. Location, Location, Location: How to select the best site for your business. Oasis Press/PSI Research, 1996.

Sen, Boudhayan, Jiwoong Shin, and K. Sudhir."Demand externalities from co-location: Evidence from a Natural Experiment." Yale University, mimeo (2011).

Terza, Joseph V., Anirban Basu, and Paul J. Rathouz. "Two-stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Eodeling." Journal of Health Economics 27, no. 3 (2008): 531-543.

Terza, Joseph V. "Correct Standard Errors for Multi-Stage Regression-Based Estimators: A Practitioner's Guide with Illustrations," Unpublished manuscript, Department of Economics, University of North Carolina at Greensboro (2012a)

Vitorino, Maria Ana. "Empirical entry games with complementarities: An application to the shopping center industry." Journal of Marketing Research 49, no. 2 (2012): 175-191.

Wooldridge, Jeffrey M. "Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables." Journal of Econometrics 182 (2014):226-234.

FOOTNOTES

[1] The retailing literature has also referred to multi-purpose shopping by consumers as ‘destination shopping’ and ‘trip chaining’.

[2] We use agglomeration effect and spillover effect interchangeably.

[3] Our study is more relevant to the demand based causal inference methods of Sen et al. (2011). There is another stream of research that employs structural supply-side models to infer agglomeration that is less relevant to the approach in our study.

[4] We present a detailed list of retail brands included in for each category of stores in the appendix.

[5] In the appendix, we present detailed estimation results that show parameter estimates are consistent regardless of which baseline stores are chosen by including household fixed effects.

[6] We discuss the specific instrumental variables that we use in our estimation

[7] Note that the figures of toy, electronics, and big box stores are combined for the sake of space as they show very similar patterns of seasonality.

[8] Since we wanted to have sufficient observations per grocery store to measure the store fixed effect, we sampled areas with sufficient numbers of households. Therefore, we excluded ZIP codes that had less than ten households in the panel. In addition, we excluded samples of households that had only one grocery store in their consideration set because we applied conditional logit regression which requires at least two grocery stores per consideration set.

TABLES

Table1: Descriptive Statistics for Household Grocery Shopping Behavior

Variable	Mean	Std. Deviation
No. visits per day	0.25	0.19
Total expenditure (dollars) per visit	39.16	25.77
Distance from home	3.50	6.07

Table 2: Endogenous Equation Parameter Estimates (Probit Model)

	Toy	**Electronic**	**Big Box**	**Home**	**Office**
$Intercept$	-3.161 (0.010)	-3.005 (0.009)	-2.953 (0.025)	-3.033 (0.013)	-3.008 (0.013)
$Seasonal\_Demand_{it}$	1.049 (0.028)	1.231 (0.037)	0.492 (0.013)	0.729 (0.010)	1.271 (0.024)
$Non\_Seasonal\_Demand_{it}$	1.137 (0.035)	1.322 (0.015)	0.528 (0.008)	0.759 (0.015)	1.453 (0.043)
$Pre\_Three\_Wks_{it}$	0.339 (0.012)	0.254 (0.024)	0.057 (0.012)	0.179 (0.011)	0.353 (0.020)
D_{ij}^{k}	-0.145 (0.007)	-0.170 (0.006)	-0.058 (0.009)	-0.151 (0.006)	-0.178 (0.009)

*standard error in parenthesis

Table 3: Validity Check on Excluded Variables

Mcfadden's Psuedo R-squared	Toy	Electronic	Big Box	Home	Office
Without excluded variables	<0.00	0.05	<0.00	0.01	<0.00
With excluded variables	0.41	0.38	0.40	0.44	0.37

Table 4: Parameter Estimates on the Main Equation

	Mean	S.E.	2.50%	50%	97.50%
α_{0i}	-0.053	0.075	-0.194	-0.052	0.095
D_{ij}	-0.275	0.033	-0.340	-0.276	-0.211
I_{ijt}^{k}	0.165	0.055	0.056	0.167	0.268
$\widehat{\tau_{ijt}^{k}}$	0.140	0.043	0.056	0.140	0.222
sf_j	√	√	√	√	√

Table 5: Agglomeration Effect Depending on Popularity of Stores

	No Non-Grocery Visit (a)	Non-Grocery Visit <1mile (b)	% Increase in Market Share (b-a)/b
Top 10% of Store Fixed Effect Stores	33.21%	37.39%	12.58%
Top 25% of Store Fixed Effect Stores	29.47%	33.29%	12.96%
Average Across All Stores	24.05%	27.18%	13.01%
Bottom 25% of Store Fixed Effect Stores	15.72%	17.98%	14.37%
Bottom 10% of Store Fixed Effect Stores	13.04%	14.99%	14.95%

Table 6: Model Estimation Varying the Definition of Agglomeration

	Maximum distance between non-grocery store and grocery store for proximity		
	1 mile	2 mile	3 mile
D_{ij}	-0.275 (0.033)	-0.288 (0.032)	-0.296 (0.031)
$\sum_k I_{ijt}^k$	0.165 (0.055)	0.135 (0.076)	0.075 (0.063)

Table 7: Model Estimation with Multiple Distance Band

Parameter	Estimates
D_{ij}	-0.298 (0.031)
$I(0 < Visit\_Dist_{jk} < 1)$	0.234 (0.057)
$I(1 < Visit\_Dist_{jk} < 2)$	-0.047 (0.058)

Table 8: Parameter Estimates for Different Non-Grocery Product Categories

	Mean	S.E.	2.50%	50%	97.50%
D_{ij}	-0.289	0.031	-0.351	-0.290	-0.230
I_{ijt}^{Toy}	0.587	0.103	0.401	0.600	0.761
$I_{ijt}^{Electronic}$	0.069	0.071	-0.086	0.074	0.201
$I_{ijt}^{Big\ Box}$	0.379	0.066	0.235	0.381	0.502
I_{ijt}^{Home}	0.404	0.074	0.249	0.403	0.549
I_{ijt}^{Office}	0.066	0.136	-0.171	0.060	0.311

Table 9: Agglomeration Effect Depending on Product Category

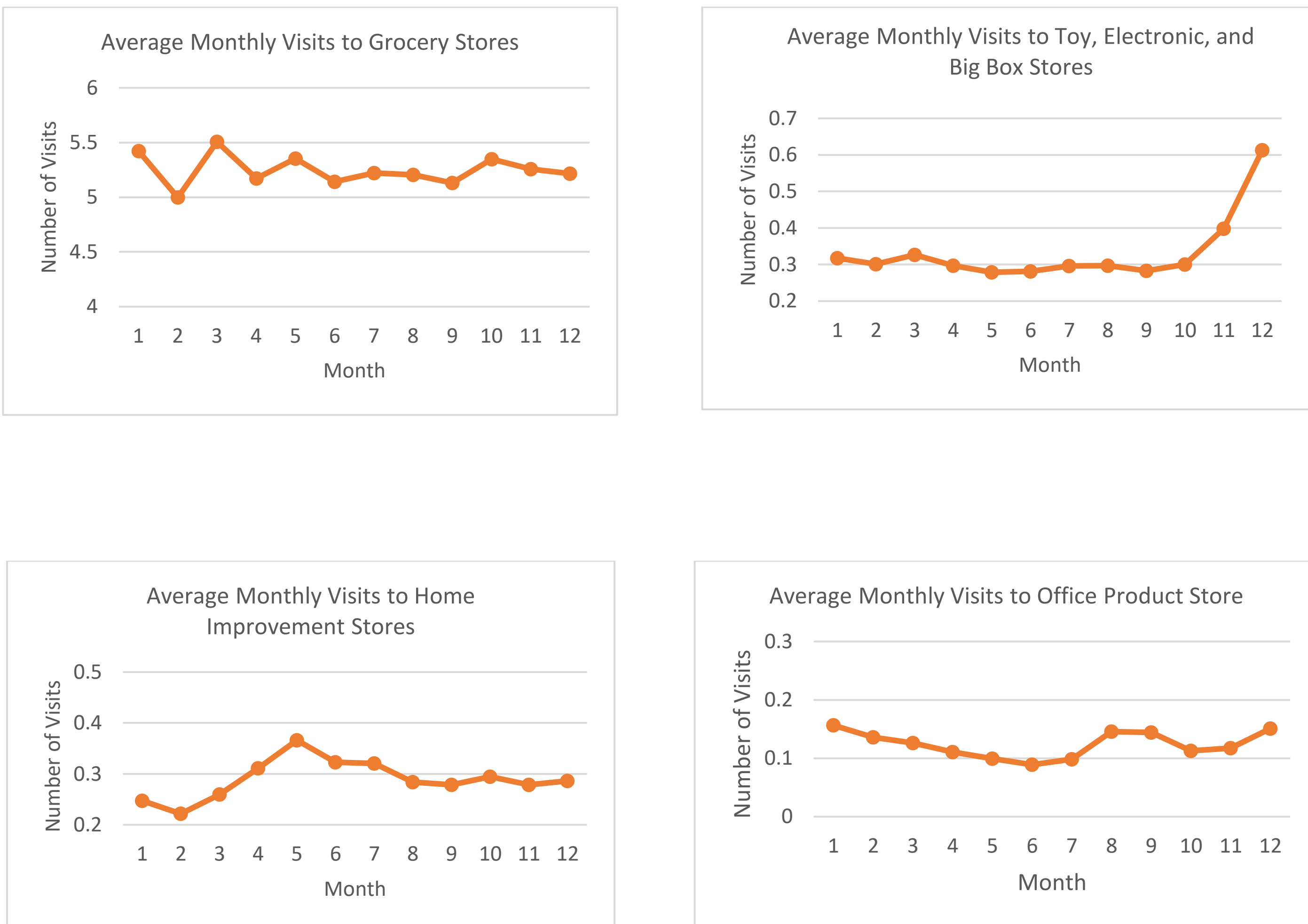
	No Non-Grocery Visit	Non-Grocery Visit <1mile	Percentage Increase in Choice Probability
Toy	24.05%	33.30%	38.46%*
Big Box	24.05%	30.11%	25.19%
Home Improvement	24.05%	30.33%	26.11%

Table 10. Agglomeration Effect Calculated Incorporating Traffic per Non-grocery Store

Store type	Spillover Rate	Agglomeration Effect in percentage
Toy	9.25%	1.39%
Big-box	6.06%	8.36%
Home	6.28%	2.58%

FIGURE

Figure 1: Seasonal Demand across Different Store Types



Appendix A

The Hierarchical Model and the Markov Chain Monte Carlo Algorithm

In this section, we provide detailed procedure of Hierarchical Bayesian Estimation. The first stage estimation is probit regression in which we used Gibbs sampling approach. The second stage estimation is multinomial logit regression where we applied Gaussian Random Walk Metropolis method.

Probit model estimation

First stage prior for β_i^k

- $\beta_i^k \sim N(\overline{\beta^k}, V_\beta)$

Second stage prior

- $V_\beta \sim IW(\nu, V)$
- $\overline{\beta^k} | V_\beta \sim N(\Delta, V_\beta A^{-1})$

where $\nu = m + 3, V = \nu \times 0.1 I_m, \Delta = 0, \mathrm{A} = 0.01$, and m is the dimension of the β_i^k

We use Gibbs sampler as follows.

- $U_{ijt}^k | \beta_i^k, Z_{ijt}^k, I_{ijt}^k$

Draw U_{ijt}^k from $trun_{(a_{ijt}, b_{ijt})}(-{Z_{ijt}^k}' \beta_i^k, 1)$ with $a_{ijt} = 0$ if $I_{ijt}^k = 1$, $a_{ijt} = -\infty$ otherwise, and $b_{ijt} = 0$ if $I_{ijt}^k = 0$, $b_{ijt} = -\infty$ otherwise.

- $\beta_i^k | U_{ij}^k, Z_{ij}^k, \overline{\beta^k}, V_\beta$

where $U_{ij}^k = \begin{pmatrix} U_{ij1}^k \\ U_{ij2}^k \\ \vdots \\ U_{ijt}^k \end{pmatrix}$, $Z_{ij}^k = \begin{pmatrix} {Z_{ij1}^k}' \\ {Z_{ij2}^k}' \\ \vdots \\ {Z_{ijt}^k}' \end{pmatrix}$ and $I_{ij}^k = \begin{pmatrix} I_{ij1}^k \\ I_{ij2}^k \\ \vdots \\ I_{ijt}^k \end{pmatrix}$

Draw β_i^k by using standard normal theory

$$\beta_i^k | U_{ij}^k, Z_{ij}^k \sim N(\widetilde{\beta_i^k}, \left({Z_{ij}^k}' Z_{ij}^k + V_\beta^{-1}\right)^{-1}); \widetilde{\beta_i^k} = \left({Z_{ij}^k}' Z_{ij}^k + V_\beta^{-1}\right)^{-1} ({Z_{ij}^k}' U_{ij}^k + V_\beta^{-1} \overline{\beta^k})$$

- $\overline{\beta^k} | \beta^k, V_\beta, A, \Delta$

Draw $\overline{\beta^k}$ from

$N(\widetilde{\beta^k}', V_\beta(1'1 + A)^{-1})$ where

$\widetilde{\beta^k} = (1'1 + A)^{-1}(1'\beta^k + A\Delta)$

and

$$\beta^k = \begin{pmatrix} \beta_1^{k\prime} \\ \beta_2^{k\prime} \\ \vdots \\ \beta_n^{k\prime} \end{pmatrix}$$

- $V_\beta | \beta^k, \nu, V, \Delta, A$

Draw V_β from $IW(\nu + n, V + S_\beta)$ where

$S_\beta = E_\beta' E_\beta$

and

$E_\beta = (\beta^k - 1\widetilde{\beta^k})' + (\widetilde{\beta^k}' - \Delta)' A(\widetilde{\beta^k}' - \Delta)$

Multinomial logit model estimation

First stage prior for α_i and sf_j

- $\alpha_i \sim N(\bar{\alpha}, V_\alpha)$
- $sf_j \sim N(\overline{sf}, V_{sf})$

Second stage prior α_i

- $V_\alpha \sim IW(\varpi, \mathrm{W})$
- $\bar{\alpha} | V_\alpha \sim N(\theta, V_\alpha B^{-1})$

where $\varpi = p + 3, \mathrm{W} = \varpi \times 0.1 I_p, \theta = 0, B = 0.01$, and p is the dimension of the α_i

Note that there is no second stage prior for the store fixed effects sf_j because sf_j are estimated only at the store level. Thus, they do not have a hierarchical structure.

The basic procedure is as follows.

- Draw α_i and sf_j using Gaussian Random-Walk Metropolis method

Random-Walk Metropolis with increments having covariances $s_{\alpha}^{2}V_{\alpha}$ and $s_{sf}^{2}V_{sf}$ relatively where $s_{\alpha}^{2} = 0.7$ and $s_{sf}^{2} = 0.008$ are scaling constants. We chose scaling constants close to $2.93/\sqrt{d}$ where d is the dimension of the parameters under interest following the guideline of Roberts and Rosenthal (2001).

- $\bar{\alpha}\,|\alpha, V_{\alpha}, B, \theta$

Draw $\bar{\alpha}$ from

$N(\tilde{\alpha}', V_{\alpha}(1'1 + B)^{-1})$ where

$\tilde{\alpha} = (1'1 + B)^{-1}(1'\alpha + B\,\theta)$

and

$$\alpha = \begin{pmatrix} {\alpha_1}' \\ {\alpha_2}' \\ \vdots \\ {\alpha_n}' \end{pmatrix}$$

- $V_{\alpha}|\alpha, \varpi, \mathrm{W}, B, \theta$

Draw V_{α} from $IW(\varpi + n, \mathrm{W} + S_{\alpha})$ where

$S_{\alpha} = {E_{\alpha}}'E_{\alpha}$

and

$E_{\alpha} = (\alpha - 1\tilde{\alpha})' + (\tilde{\alpha}' - \theta)'B(\tilde{\alpha}' - \theta)$

Appendix B

List of the Retail Brands under each Store Category

Table B shows the specific names of the retail brands we considered for our study.

Table B: List of the Retail Brands under each Store Category

Store Type	**Retail brands**
Toy Store	Toys R Us, Babies R Us, Kaybee Toy, Kids R Us, Zany Brainy
Electronic Store	Best Buy, Circuit City, Computer City, Frys Electronics, PC Connection, PC Zone, Radio Shack, Comp USA
Big-box Store	Kmart Big, Target, Walmart
Home Improvement Store	Home Depot, Lowes, Menards, Builders Square
Office Product Store	Staples, Office Depot, Office Max

Appendix C

This section provide with tables related to the combination of non-grocery stores near grocery stores.

Table C1: Number of grocery stores having each combination of non-grocery stores within a one-mile distance

Non-grocery Store Combination	Number of Grocery Store	Percentage
Toy	13	1.3%
Electronic	197	20%
Big Box	35	3.6%
Home	29	2.9%
Office	22	2.2%
Toy+Electronic	15	1.5%
Toy+Big Box	8	0.8%
Toy+Home	3	0.3%
Toy+Office	2	0.2%
Electronic+Big Box	22	2.2%
Electronic+Home	58	5.9%
Electronic+Office	10	1%
Big Box+Home	9	1%
Big Box+Office	11	1%
Office+Home	7	0.7%
Toy+Electronic+Big Box	26	2.6%
Toy+Electronic+Home	42	4.3%
Toy+Electronic+Office	3	0.3%
Electronic+Big Box+Home	14	1.4%
Electronic+Big Box+Office	7	0.7%
Big Box+Home+Office	2	0.2%
Toy+Big Box+Home	2	0.2%
Toy+Big Box+Office	3	0.3%
Toy+Home+Office	2	0.2%
Electronic+Home+Office	8	0.8%
Toy+Electronic+Big Box+Home	54	5.5%
Toy+Electronic+Big Box+Office	5	0.5%
Toy+Big Box+Office+Home	3	0.3%
Electronic+Big Box+Office+Home	6	0.6%
Toy+Electronic+Home+Office	11	1.1%
Toy+Electronic+Big Box +Office +Home	24	2.4%

NONE	314	32%
TOTAL	967	100%

Table C2: Percentage of single vs. multi purpose shopping incidences per Non-grocery Store Combination

Non-grocery Store Combination	Percentage of Single Purpose Shopping	Percentage of Multi-Purpose Shopping
Toy	0.29%	0.21%
Electronic	12.66%	8.27%
Big Box	3.24%	1.55%
Home	1.77%	1.19%
Office	2.04%	1.09%
Toy+Electronic	0.72%	0.42%
Toy+Big Box	0.57%	0.27%
Toy+Home	0.10%	0.14%
Toy+Office	0.38%	0.21%
Electronic+Big Box	2.36%	1.35%
Electronic+Home	2.94%	2.07%
Electronic+Office	0.83%	0.54%
Big Box+Home	0.56%	0.33%
Big Box+Office	0.86%	0.42%
Office+Home	0.89%	0.38%
Toy+Electronic+Big Box	1.41%	1.00%
Toy+Electronic+Home	2.16%	1.54%
Toy+Electronic+Office	0.22%	0.14%
Electronic+Big Box+Home	1.21%	0.59%
Electronic+Big Box+Office	0.64%	0.53%
Big Box+Home+Office	0.21%	0.09%
Toy+Big Box+Home	0.13%	0.06%
Toy+Big Box+Office	0.24%	0.13%
Toy+Home+Office	0.05%	0.03%
Electronic+Home+Office	0.31%	0.20%
Toy+Electronic+Big Box+Home	3.99%	2.59%
Toy+Electronic+Big Box+Office	0.29%	0.15%
Toy+Big Box+Office+Home	0.33%	0.19%
Electronic+Big Box+Office+Home	0.66%	0.29%
Toy+Electronic+Home+Office	0.15%	0.16%

Toy+Electronic+Big Box +Office +Home	1.60%	1.11%
NONE	18.54%	10.40%
Total	62%	38%

Appendix D

Comparison of estimation results with household fixed effects vs. without household fixed effects

In this section we show the estimation results to show how including household fixed effects to the model help us getting more consistent parameter estimates compared to the model without the effects.

In order to test the benefit, we estimated two types of model (i.e., with household fixed effect vs. without household fixed effect) on three data set ups differs in terms of baseline stores. Specifically, for data1 we chose baseline store to be the least popular store by the household within his/her consideration set, for data3 we chose baseline store that is the most popular by the household, and for data2 we chose a random baseline store within his/her consideration set.

Table D1 shows the parameter estimates under the model without household fixed effects while Table D2 shows the result under the model with household fixed effect. As can be seen from the result, parameter estimates are only consistent under the model with household fixed effect.

Table D1: Parameter estimates under the model without household fixed effects

	Data1	Data2	Data3
D_{ij}	-0.249 (0.033)	-0.272 (0.031)	-0.132 (0.029)
I_{ijt}^{k}	-0.306 (0.050)	0.192 (0.057)	0.796 (0.053)
$\widehat{\tau_{ijt}^{k}}$	0.573 (0.029)	0.126 (0.026)	-0.493 (0.026)
sf_j	√	√	√

Table D2: Parameter estimates under the model with household fixed effect

	Data1	Data2	Data3
α_{0i}	0.838 (0.081)	-0.053 (0.075)	-1.211 (0.062)
D_{ij}	-0.258 (0.032)	-0.275 (0.033)	-0.157 (0.031)
I_{ijt}^{k}	0.196 (0.067)	0.165 (0.055)	0.212 (0.051)
$\widehat{\tau_{ijt}^{k}}$	0.128 (0.045)	0.140 (0.043)	0.092 (0.036)
sf_j	√	√	√