A Gravity Model for Retail Store Entry

Boram Lim
Hanyang University
Harsha Kamatham
University of Manitoba

Cite as:
Lim Boram, Kamatham Harsha (2023), A Gravity Model for Retail Store Entry. *Proceedings of the European Marketing Academy*, (117158)

Paper from the EMAC Regional Conference, Athens, Greece, September 27-29, 2023

A Gravity Model for Retail Store Entry

Abstract

We build an empirical framework of retail store location choice for a retail chain manager based on the law of retail gravitation, the foundational theory of store choice. We estimate the model using household loyalty-card data from a multichannel retail chain located in metropolitan areas. Along with size of retail outlet, prices, promotion, and distance to a retail store, we particularly measure and quantify the effects of (i) the distance of a retail store from households' residence, and (ii) the actual weight of the shopping basket on the store-choice decision. In particular, to our best knowledge, this paper is the first in the literature to quantify the effect of actual weight of the shopping basket à la the law of retail gravitation. We also find that when the weight of a basket increases by one pound, the likelihood of choosing the nearest brick-and-mortar store increases by 1.43%. In a counterfactual analysis, we identify and compare the profitability of candidate locations for a potential new chain store.

Keywords: store choice,law of retail gravitation

1 Introduction

As more and more grocery retailers are becoming multi-channel retailers, the number of brick-and-mortar stores that also serve online orders and provide in-store pickup services or delivery services is increasing (Dawes & Nenycz-Thiel, 2014; Neslin & Shankar, 2009). When a manager of a multi-channel retail chain seeks to open a new brick-and-mortar store that also serves online orders (multi-channel store hereafter), assessing the demand of potential locations for opening a multi-channel store is more complex than assessing the demand of potential locations for opening a traditional brick-and-mortar store. First, the extent to which one multi-channel can cover the geographical area is unclear. On the one hand, because customers do not have travel costs when they make online purchases, the geographical reach of the multi-channel store could be anywhere the store can provide delivery services. On the other hand, the presence of a nearby retail outlet increases consumer consideration of the brand by creating top-of-mind awareness (Shriver & Bollinger, 2015). In other words, if a physical multi-channel store chain is not located nearby, customers are less likely to use the online channel of the store chain. Second, though previous literature (e.g., Bell, Ho, & Tang, 1998; **?**) shows how the pricing format of outlets, assortment size, and distance to store affect customers' store-choice behavior, one must understand why customers use the online channel[1] of the retail chain or one of the nearby offline outlets when the online channel is another alternative option with other offline stores. Thus, the multi-channel retailer has difficulty predicting the demand for each of its online and offline stores. Therefore, the multi-channel store's location decision requires the understanding of customers' store-choice behavior while incorporating the existence of an online store. However, previous research on store choice focuses only on customers' physical-store-choice behavior (e.g., Bell et al., 1998; Briesch, Chintagunta, & Fox, 2009; **?**), and studies on customer channel choice behavior (e.g. Chintagunta, Chu, & Cebollada, 2012) are based on the assumption that customers already choose the store chain before deciding a shopping channel. Furthermore, empirical progress on identifying a set of retail locations that would produce maximum revenue or profit has been limited, though the insight from early analytical models can help retailers identify the attributes of attractive locations (Glaeser, Fisher, & Su, 2019). Presumably, modeling customer store-choice behavior requires extensive data such as all

[1] We use "online channel" ("offline channel") and "online store" ("offline store") interchangeably.

the store-choice records of households from a certain geographical area with the information of all competing chains' pricing formats, promotions, and product assortments. In this paper, we build a store-choice model by constructing individualized store choice sets consisting of nearby offline stores and an online store with unique data sets from a unique retail environment. In a counterfactual analysis, we select potential locations for opening a new store and compare the profitability of candidate locations for a new chain store.

2 Data

2.1 Focal Grocery Chain

Our data is collected from a major grocery retail chain in Korea. It has a prominent presence throughout South Korea. It has around 300 physical stores and an online store. The centralized online store, linked to all the retailer's physical stores, allows customers to place orders on the website, with the orders being filled and delivered by the nearest physical store.

The online store and offline stores provide the same range of products, aside from a few categories such as alcohol, cigarettes, and a few ready-to-eat food products that are only available through the offline channel. The retailer is a Hi-Lo chain and it practices uniform pricing and runs the chainwide promotions. In addition, prices are identical across shopping channels. However, the retailer has different promotions between online and offline.[2]

2.2 Household Panel Data

We obtained the complete shopping records of 6000 households between September 2013 and August 2014.[3] For each shopping trip undertaken by households in our sample, we observe the time of the shopping trip, the items purchased, the purchase price, and the channel used. After removing households with missing demographic information and households that had too few shopping trips, we are left with 5,337 households. As mentioned earlier, since our study aims to investigate customers who do grocery shopping on foot, we select stores located in city with more than 500,000 population and select customers living within a 1.5 mile radius from each of offline store.

[2]There are rarely channel specific promotions. The price differences between the channels caused by the promotions are considered in the analysis

[3]In our sample, we have a pre-analysis period (Sep 2013 - Dec 2013) and an analysis period (Jan 2014 - Aug 2014).

Table 1: Household Demographics and Characteristics of Focal chain store

	Sample for analysis (2605)	Entire Sample (5337)
Characteristics of Households		
Age (loyalty card holder in a household)	42.79 (7.99)	41.90 (7.86)
Live in apartment	60.70%	57.15%
Marriage (married =1)	54.09%	52.78%
Gender (Female = 1)	86.36%	84.04%
Percentage of Offline trips	63.79% (38.2%)	62.00% (39.46%)
Number of Offline only households	930	1950
Number of Online only households	209	570
Number of mixed channel households	1466	2817
Characteristics of offline store of focal chain		
Number of offline stores	190	205
Number of HH within a 0.19 mile (300m) radius of store	2943.11 (1287.5)	2940.59 (1256.44)
Number of HH within a 0.31 mile (500m) radius of store	6907.67 (3084.67)	6897.91 (3024.34)
Number of HH within a 0.43 mile (700m) radius of store	11559.46 (5630.3)	11557.24 (5554.56)
Number of HH within a 0.62 mile (1km) radius of store	19685.3 (10850.29)	19670.07 (10757.19)
Percentage of HH living in APT within a 0.19 mile (300m) radius of store	72% (30%)	73% (30%)
Percentage of HH living in APT within a 0.31 mile (500m) radius of store	69% (27%)	70% (27%)
Percentage of HH living in APT within a 0.43 mile (700m) radius of store	67% (25%)	67% (25%)
Percentage of HH living in APT within a 0.62 mile (1km) radius of store	64% (23%)	65% (23%)

After removing customers who report wrong address information, we use 2605 households out of 5337 households for the analysis. Table 1 presents major demographics and store characteristics regarding household distribution near our focal chain stores.

3 Model

Our objective is to measure the gravity effect of basket weight and location and quantify the effects on household store choice behavior. In accordance with this objective, we model store choice model at the level of individual shopping trip occasions. The decision on which store she wants to shop depends on distance to stores, price, promotion, expected weight of basket and store size. Inspired by the gravity equation, we specify the utility that a household i derives choosing store j at time t for the composition of basket x_{it} is:

$$
\begin{aligned}
U_{ijt} = {} & \beta_0 + \beta_1 P(x_{it}; j) + \beta_2 W(x_{it}; j) + \beta_3 D_{ij} + \beta_4 D_{ij}^2 \\
& + \beta_5 D_{ij} * W(x_{it}) + \beta_6 P(x_{it}; j) * W(x_{it}) \\
& + \beta_7 Wnd_t + S_{ij}\theta + HH_i\delta + \epsilon_{ijt}
\end{aligned} \tag{1}
$$

β_0 fixed cost/benefit of shopping somewhere, P_{ijt} is the price index of basket x_{it} in store j. We calculate the price indices based on price information from two public institutions in South Korea. The institutions publish the prices of 400 necessities weekly and 400 agricultural and marine products daily for each of major grocery chain. We assume that households choose a store to shop based on their expected price across nearby stores and the price information from the institutions can represent the households expectation across the nearby grocery chains. Because the products that the institutions choose to publish the price information are most frequently purchased items in general we assume the expected prices differences across chains can be captured by prices of those products.

$W(x_{it}; j)$ is the basket weight, D_{ij} distance to store j. If a store j is the focal chain's online store we assume that D_{ij} and $W(x_{it}; j)$ are zero, because all the online order in our sample are delivered households do not have utility/disutility from basket weight and distance for online orders[4]. That is, basket weight, $W(x_{it}; j)$, is the same for all individual i's nearby offline stores and 0 for online store of the focal grocery chain. Thus, the coefficient, β_2 captures the effect of weight of basket on choosing the online channel of our focal grocery chain. We include D_{ij} and D_{ij}^2 in the utility equation. As shown in Figure **??**, the channel usage pattern between online and offline stores of the focal chain show quite differently and the relationship between the distance to focal store chain and the expenditure through online channel of the focal chain shows the inverted U-shape. Thus, we assume the relationship between the store choice decision and the distance is non-linear. Wnd_t is the weekend dummy (1 if day t is a weekend, and 0 otherwise)

We also include the information of store size, S_{ij}, which consists of three dummy variables: small size, medium size, and online store (baseline is the large size store). Previous literature show quite different result on the effect of store size on store choice decisions. Fox, Montgomery, and Lodish (2004) and Briesch et al. (2009) show the positive relationship between the assortment

[4] In the online store, there is no pick-up option and deliveries from offline shopping are negligible in the data.

size and the store choice decisions. On the other hand, Boatwright and Nunes (2001) shows the negative relationship between assortment size and category sales. So, we include the size of stores as a characteristics of store in store choice decision. We group sizes of stores based on the total area of stores by assuming that stores with larger area are have larger product assortments. HH_i is the households' demographics including age, whether living in APT, gender and marital status. ϵ_{ijt} is an idiosyncratic utility shock and we assume it follows Type I extreme value distribution.

$V_{ijt}(x_{it})$ is the utility from purchasing basket x_{it}. Let $\mathcal{J}_i = \{0, 1, 2, ..., J_i\}$ denote the set of nearby supermarkets to households i's home. 0 here denotes the "outside option," carrying the basket needs to tomorrow. Households problem can be formulated in the following two stages:

1. Choose whether to shop today. If the outside option is chosen, the basket need carries over, i.e., $x_{it+1} = x_{it} \cup \tilde{x_{it}}$, where $\tilde{x_{it}}$ is the "flow of need."
2. Choose the store that maximize U_{ijt} from $\{\mathcal{J}_i \setminus 0\}$.

The challenge we face here is that we only observe the basket compositions for the choices of offline and online stores at the focal chain. We construct the flow of need with only the basket composition we observe. When we observe the consecutive shopping trips at time at time $t-n$ and t with basket compositions, x_{it-n} and x_{it} respectively (n days elapsed between two observed shopping at the focal chain), the flow of need between time $t-n$ and t is interpolated by assuming the need of x_{it} is proportionally increased by interpurchase time from zero. That is, the amount of need at time $t-n+1$ is the x_{it}/n and so on.

It is indeed possible that the observed basket weights are not exogenous variable. The unobserved factors that affect the store choice decision could also affect the weight of basket. We employ a control function approach Petrin and Train (2010) to account for potential endogeneity of basket weight. The idea behind the control function correction is to derive a proxy variable that conditions on the part of $W(x_{it})$ that depends on ϵ_{ijt}. The control function approach seeks to recover the unobserved portion in a first step and then to insert it into the utility function (1). In the first step, we regress the potentially endogenous basket weight variable on a number of instruments as well as on exogenous variables of the utility equation:

$$W(x_{it}; j) = \delta J_{ijt} + \gamma X_{ijt} + \lambda_i + \eta_{ijt} \tag{2}$$

where J_{ijt} and X_{ijt} are vectors of instrument variables and observed characteristics in the utility equation respectively, $\lambda 1_i$ is a individual fixed effects. The error-term,η_{ijt}, contains the unobserved factors that are not captured by observed store characteristics. In the second step, the resudual retained from (2) is plugged into the utility function:

$$\begin{aligned} U_{ijt} = \beta_0 + \beta_1 P(x_{it}; j) + \beta_2 W(x_{it}; j) + \beta_3 D_{ij} + \beta_4 D_{ij}^2 \\ + \beta_5 D_{ij} * W(x_{it}) + \beta_6 P(x_{it}; j) * W(x_{it}) \\ + \beta_7 Wnd_t + S_{ij}\theta + HH_i\delta \\ + \tau\widehat{\eta}_{ijt} + \bar{\epsilon}_{ijt} \end{aligned} \tag{3}$$

Assuming that $\bar{\epsilon}_{ijt}$ is iid Type I extreme value distributed, we can write the conditional probability in terms of choice-specific expected value function

$$Pr(i \text{ chooses } j | x_{it}; \beta) = \frac{exp(V_{ijt})}{\sum_{k \in \mathcal{J}_i} exp(V_{ikt})} \tag{4}$$

4 Result

We report the parameter estimates for the store choice component of the model in Table 2. We find that the basket weight parameter is significantly negative. The coefficient for the basket weight captures the effect of basket weight on choosing store between online store and other offline stores. It suggests that households are likely to choose online store as their weight of baskets increase (coefficient = -0.495; std-error = 0.008). The estimate for price index is significantly negative (coefficient = -0.532; std-error = 0.022), as we expect.

We include square term of distance to store in the store choice utility model. We find that the distance to store is non-linearly related to the store choice decision (coefficient for the square term of distance = 1.497; std-error = 0.023). Previous literature that consider the store choice behavior within only offline stores assume the linear relationship between the distance to a store and the store choice decision. However, when an online store exists under customers' choice set, the disutility from the distance is not simply linear. We find that as the distance to offline store increases households tend to use online store. We find this pattern in the relationship between the distance to focal store chain and the probability of choosing focal chain's shopping channel. It shows that as a household who uses both online and offline stores lives further from an offline store the likelihood of choosing the offline store decreases. But the pattern in online store is different. They are more likely to choose offline store if they live very close to the offline store (within 0.1 miles), but if the distance to the offline store is more than 0.5 miles, the likelihood of choosing the online store is nearly constant regardless of the distance to the offline store. In addition, households using only offline store have a greater utility reduction with distance compared to households using both online and offline (on/off shopper). In addition, the negative impact of the distance on offline store choice is stronger for offline-only shoppers than on/off shoppers.

Table 2: Model Estimates

	Estimate	SE
Intercept	0.540	0.035
Basket weight	-0.495	0.008
Price indices	-0.532	0.022
Distance to store	-5.295	0.031
Distance to store^2	1.497	0.023
Store size - Online	-2.119	0.019
Store size - Small	-0.728	0.009
Store size - Med	0.333	0.007
Basket weight * Distance to store	-0.015	0.001
Basket weight * Price indices	0.379	0.009
Age	0.010	0.001
Live APT	0.041	0.009
Gender - female	0.021	0.012
Marriage	0.077	0.008
Weekend Shopping	-0.462	0.018
CF	5.446	0.057

We examine the relationship between a distance to focal chain's nearest offline store and the likelihood of choosing a focal chain's channel by basket weight, and it shows that as the weight of shopping basket increases the positive relationship between the distance and the likelihood of choosing online channel is stronger. When the average basket weight is smaller than 5 lbs, households are equally likely to choose online channel and offline channel when the distance to nearest focal chain's offline store is 0.49 mile (intersection between two lines in Figure 1-(a)), when the basket weight is between 5 lbs and 15 lbs they are equally likely to choose both channels at 0.24 mile (Figure 1-(b)). When the basket weight is heavier than 25 lbs, they prefer online regardless of their locations. This is a very important finding customer store modeling. So far, travel cost due

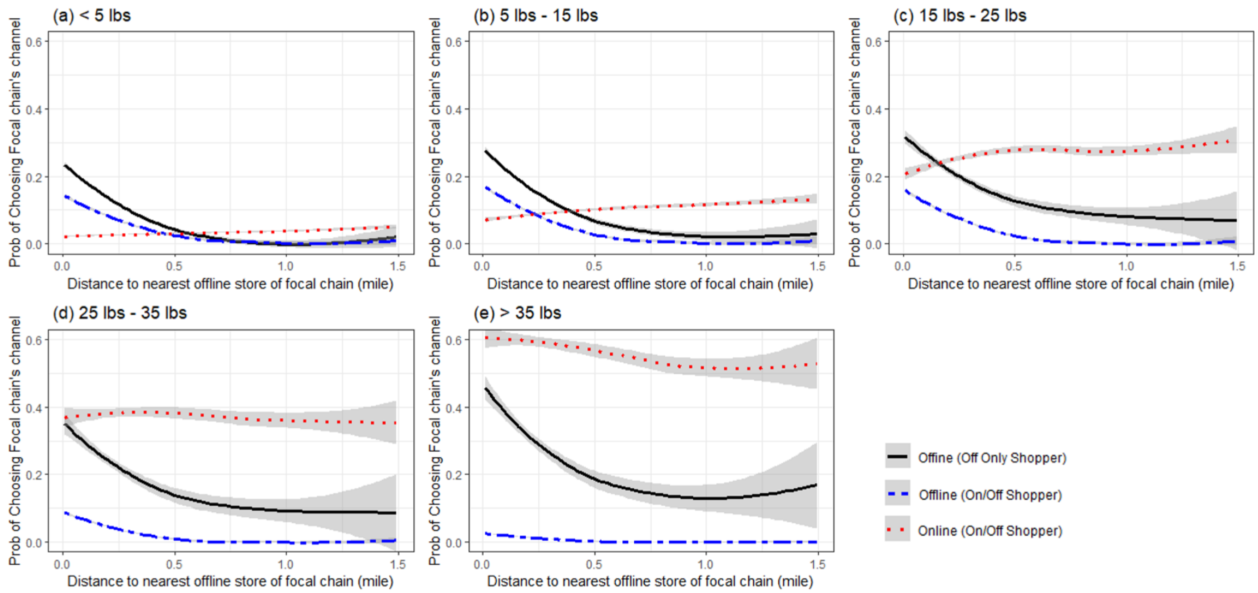

Figure 1: Expected Probability of Choosing a Store by Basket Weight

to the distance to a store has been assumed the fixed cost for shopping in literature (e.g. Bell et al., 1998; Briesch et al., 2009) employing the framework that customers choose a store to minimize the sum of fixed and variable cost of shopping. That is, previous literature assume that disutility from the distance to a store is not associated with the amount of buying items. However, our finding shows that the disutility of travel cost is associated with the size of basket and the existence of online store make strong the association because customers now use online store when they expect to purchase large size basket.

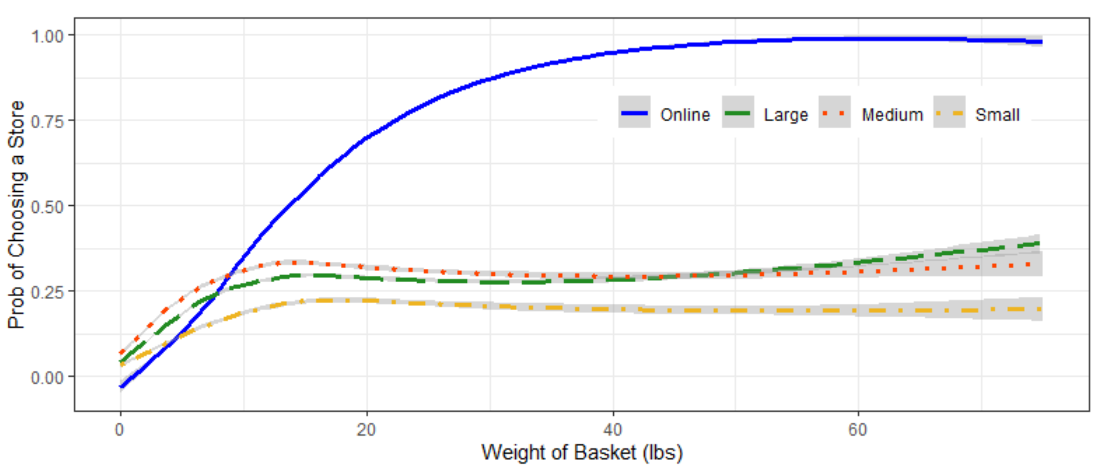

Figure 2: Probability of Choosing a Store by Store Size

To examine how the store choice decisions differ according to store size, we report the relationship between the weight of baskets and the probability of choosing a store by store sizes in Figure 2. We find that households with small basket size (less than 10 lbs) are likely to choose medium size offline store but they prefer online store with heavy basket (heavier than 10 lbs).

References

Bell, D. R., Ho, T.-H., & Tang, C. S. (1998). Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research*, *35*(3), 352–369.

Boatwright, P., & Nunes, J. C. (2001). Reducing assortment: An attribute-based approach. *Journal of marketing*, *65*(3), 50–63.

Briesch, R. A., Chintagunta, P. K., & Fox, E. J. (2009). How does assortment affect grocery store choice? *Journal of Marketing research*, *46*(2), 176–189.

Chintagunta, P. K., Chu, J., & Cebollada, J. (2012). Quantifying transaction costs in online/off-line grocery channel choice. *Marketing Science*, *31*(1), 96–114.

Dawes, J., & Nenycz-Thiel, M. (2014). Comparing retailer purchase patterns and brand metrics for in-store and online grocery purchasing. *Journal of Marketing Management*, *30*(3-4), 364–382.

Fox, E. J., Montgomery, A. L., & Lodish, L. M. (2004). Consumer shopping and spending across retail formats. *The Journal of Business*, *77*(S2), S25–S60.

Glaeser, C. K., Fisher, M., & Su, X. (2019). Optimal retail location: Empirical methodology and application to practice: Finalist–2017 m&som practice-based research competition. *Manufacturing & Service Operations Management*, *21*(1), 86–102.

Neslin, S. A., & Shankar, V. (2009). Key issues in multichannel customer management: current knowledge and future directions. *Journal of interactive marketing*, *23*(1), 70–81.

Petrin, A., & Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of marketing research*, *47*(1), 3–13.

Shriver, S., & Bollinger, B. (2015). A structural model of channel choice with implications for retail entry. *Shriver: Columbia Business School, Colubia University*.