

The Influence of Shopping Path Length on Sales Growth and Its Variance

Shunichi Ohmori

Department of Industrial and Management System Engineering,
Waseda University, Okubo 3-4-1, Shinjuku, Tokyo 169-8555, Japan
Email: ohmori0406@aoni.waseda.jp (*Corresponding Author*)

Masao Ueda

Department of Industrial and Management System Engineering,
Waseda University, Okubo 3-4-1, Shinjuku, Tokyo 169-8555, Japan
Email: m_ueda@aoni.waseda.jp

Kazuho Yoshimoto

Department of Industrial and Management System Engineering,
Waseda University, Okubo 3-4-1, Shinjuku, Tokyo 169-8555, Japan
Email: kazuho@waseda.jp

ABSTRACT

We study the influence of shopping path length on the number of items purchased. It is a well known fact in in-store shopper marketing that the longer the travel distance, the more items purchased. We analyze a shopping path data of 556 shoppers collected at a retail store in an urban area in Tokyo, Japan. We observed the fact the longer the travel distance, the greater the variance of the number of items purchased as well as its average. To explain this fact, we proposed a probability model of sequential purchase decisions using the Bernoulli process. We can estimate the number of items purchased as the binomial distribution given the shopping path. In the case study, we show how this assumption can be justified by the data from a real store.

Keywords: *shopper marketing; retail operation*

1. INTRODUCTION

In recent years, there have been tremendous changes in shoppers' behavior due to changes in technological and business models. In view of the facts of Toys 'R' Us bankruptcy in the US and Amazon's acquisition of Whole Foods, most retailers have been struggling against larger e-commerce competitors. Faced with severe competitions not only among retail industry itself but with the EC industry, the need for retailers to innovate in marketing strategy in stores has never been critical. Over the years retailers are increasing attention allocated to practice in shopper marketing. In shopper marketing, marketers frequently distinguish the planned and unplanned buying. It is believed that the majority of buying decisions are unplanned, and that shopping path length and unplanned buying are correlated. To support this belief quantitatively, the relationship between in-store travel distance and unplanned spending has been studied (Kollat and Willett 1967; Rock 1987; Granbois 1968; Hui *et al.* 2009c). Kollat and Willett (1967) claimed that "in-store stimuli usually reminds shoppers of present or

future needs rather than evoking new needs". Therefore, the longer the distance shoppers walk in the store, the more exposure to in-store stimuli to remind them of such potential needs.

In this study we extend this research on the relation between shopping path length and unplanned buying. We used shopping path data of 556 shoppers, collected in a grocery store located in an urban area in Tokyo, Japan. We observed the fact that "the longer the shopping path, the greater the variance, not just mean, of the number of the items purchased". The primal interest of this research is to analyze the reason behind this fact. We modeled a dynamical behavior of shoppers using the *visit-shop-buy* framework proposed by Hui *et al.* (2009c). We proposed a probability model of sequential purchase decisions using the Bernoulli process. We can estimate the number of items bought as the binomial distribution given the shopping path. In the case study, we show how this assumption can be justified by the data from a real store.

The reminder of the paper is as follows. In section 2, we present literature review. In section 3, we describe the data used to validate our model. In section 3, we present our proposed model. In section 4, we present the result of application.

2. LITERATURE REVIEW

This work relates to recent advancements in shopper marketing models. In-store behavior of shoppers and relationship between In-Store travel distance and unplanned spending has been studied for more than 60 years. Kollat and Willett (1967) studied customer differences in unplanned purchasing behavior. Granbois (1968) studied relation between customer response to variations in in-store factors such as price, layout, and display. Rock (1987) studied impulse buying behavior of shoppers and revealed subjective experience subjective experiences that distinguish the onset of the buying impulse, how shoppers cope with their impulsive urges to buy, and the types of negative

consequences they incur as a result of their impulsive buying. Hui et al. (2009b) decomposed the length of observed shopping path into three components: the length of the TSP-path, the additional distance because of order deviation, and the additional distance because of travel deviation. They explored the relationship between these deviations and different aspects of in-store shopping/purchase behavior. Hui et al. (2009c) proposed an integrated probability model in which a grocery path is divided into a series of three exhaustive sequential and interrelated decisions (visit, shop, and buy). Hui et al. (2013) studied the endogeneity of in-store travel distance. They used in-store radio frequency identification tracking and develop an instrumental variable approach to analyze the causal effect of in-store travel distance on unplanned purchase.

Other metrics to use are store-coverage (Granbois 1968; Hansen and Ottar Olsen 2006; Hui et al. 2013; Knox et al. 2011; Kollat and Willett 1967), the time spent in the store (Granbois 1968), and the number of aisles shopped (Granbois 1968), each of which has correlation with the unplanned buying. Clearly these measures are correlated (Sorensen 2017). For example, the proportion of a store covered may be determined by how much time the shopper has available to spend in the store or the items the shopper intends to purchase. Alternatively, the items needed may dictate how much of the store the shopper covers, which may, in turn, influence the time taken to complete a shopping trip.

Several studies have examined many other factors that drive unplanned purchases. The factors include budget deviation (Stilley et al., 2010), browsing and shopping (Hui et al. 2009), social influences (Zhang et al. 2014), and factors that influence unplanned purchases and spending (Bell et al., 2011; Hui et al. 2013; Inman, Winer, and Ferraro 2009). While this research stream has developed valuable managerial insights, most of the factors are very hard to control directly by retailers.

It is known from prior research that the shopping trips varying length depending on various factors. Those factors include shopping mission (Walters and Jamil 2003), the day of the week (Kahn and Schmittlein 1989; Tanskanen et al. 2002), store type (Sorensen 2009) and store layout (Sorensen 2016). Several papers compare shopping lengths across countries (i.e. Sorensen 2016). Sorensen (2009) studied data collected on 75,000 shoppers across a series of three stores to develop behavioral segmentation of shoppers. They developed three segmentations from mathematical clustering as quick, fill-in, stock-up. They discovered those three segments vary in shopping path length. Similarly, Inman et al. (2009) divide shoppers into three groups according to whether they visited 'all aisles', 'most aisles', 'a few aisles'. However, none of the above-cited research has focused on the fact that "the greater the distance, the greater the variance", which is the central research question in this paper.

3. DATA SET AND OBSERVATION

3.1 Data Set

The data set used in this case study is the customer survey data of shopping path in the grocery store located at Tokyo in Japan. In this survey, investigators chase shoppers and record walking path of each shopper. Operating hours of

a store is from AM 10:00 to PM 9:00. Day of the week of data collection is four successive days from Thursday to Sunday. Sample size is 556 and **Table 1** presents detail sample size for Age/Sex segments. The floor layout is presented as in **Figure 3**. The aisle space between gondolas and their intersections are divided into 170 zones and labeled as nodes. Adjacent nodes are connected by arcs.

There are 170 nodes and 471 arcs, and the entrance node is located at left-upper corner and the registers are located at nodes 50, 52, 54, 56, 58, 60, 62, 64, 66 along the left edge, each of which is connected to the exit node 171. The representation used to model the floor is as follows. Let $N = \{i | i = 0, 1, \dots, n, n+1\}$ be a set of zones with 0 being an entrance zone, $1, \dots, n$ being in-store zones, and $n+1$ being an exit zone. $S_i = \{s | s = 1, \dots, n_i\}$ denotes a set of items located in zone i , $S = S_1 \cup \dots \cup S_n$ denotes a set of all items. The shopping-path for each shopper is recorded using this zone number, e.g., $0 \rightarrow 1 \rightarrow 70 \rightarrow 69 \rightarrow 71 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow \dots \rightarrow 137 \rightarrow 58 \rightarrow 170$.

Table 1 Sample Size for Age/Sex Segments

Age	Male	Female
20-30	7	11
30-40	17	71
40-50	24	82
50-60	25	83
60-70	43	94
≥ 70	43	56
Total	159	397

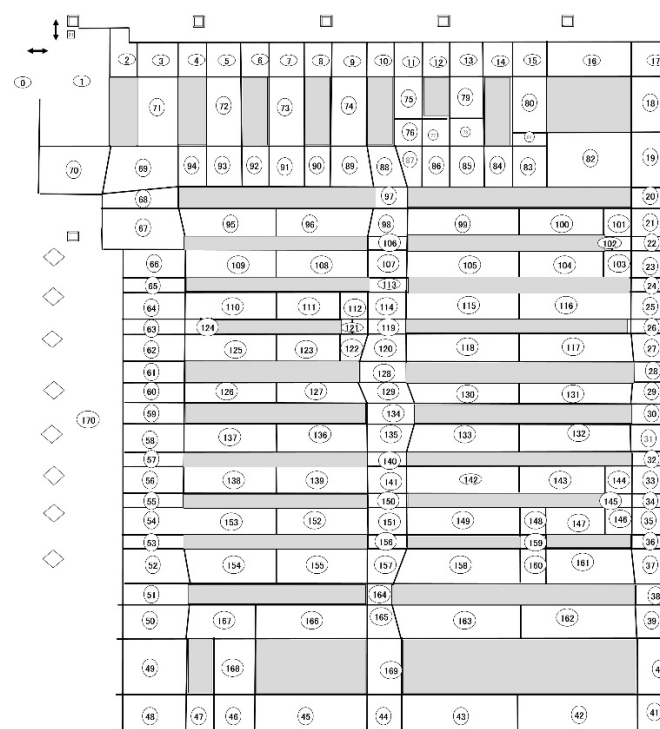


Figure 3 Floor Layout

3.2 Observation

Travel distance segmented by sex and age is presented in **Table 2**. Average distance is 207 meters for total shoppers, 180 meters for male, and 213 meters for female. We see that female walks longer distance in all age segments.

To see the relationship between the travel distance and the number of items purchased, we made the scatter diagram between the number of items purchased and distance, shopping time, the number of zones visited in Figure 2, 3, and 4 respectively. From the **Figure 2**, we see the relationship that “the longer the distance, the more items purchased”. However, we see that some of shoppers that walked long, say over 400 meters, did not spend a lot. On the other hand, there are some shoppers who bought many items even though the path-length is average, say around 300 meters.

We see the same tendency from **Figure 3** and **Figure 4**. Looking at the result, while the correlation between path-length and the number of items is confirmed, there are some shoppers that cannot be analyzed under this framework.

Especially, we realized the fact those *deviate* shoppers increase as the travel distance increase. To say that in another word, we have that “the longer the distance, the greater the variance”. This fact is merely mentioned in the previous studies. It prompted needs to segment shoppers and know which factors that affect the number of purchases for each of segments. The data set used in this case study is the customer survey data of shopping path in the grocery store located at Tokyo in Japan. In this survey, investigators chase shoppers and record walking path of each shopper.

Table 2 Travel Distance by Age/Sex Segments

Age	Average Distance			Standard Deviation		
	M	F	Total	M	F	Total
20-30	185	241	213	0	0	28
30-40	142	256	213	63	91	98
40-50	188	204	201	65	69	68
50-60	196	228	221	71	72	73
60-70	175	228	215	51	80	78
70	179	208	198	80	78	80
Total	180	217	207	73	77	78

(M: Male, F: Female)

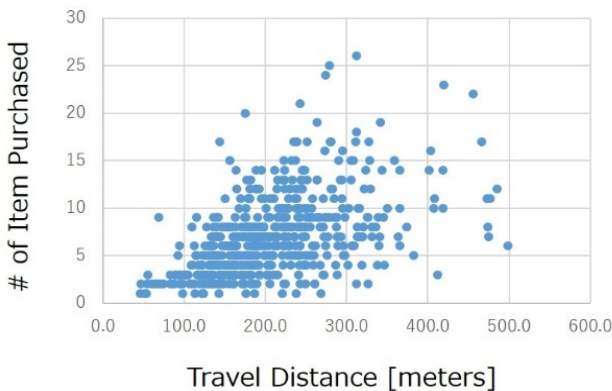


Figure 2 Scatter Diagram between Distance and # of Items Purchased



Figure 3 Scatter Diagram between Shopping Time and # of Items Purchased

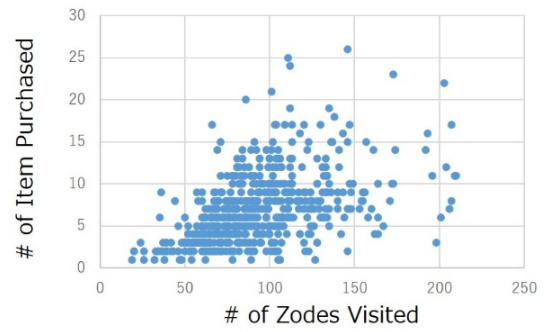


Figure 4 Scatter Diagram between # of Zones Visited and # of Items Purchased

4. PROPOSED MODEL

The proposed model divide a shopping path into a series of three sequential and interrelated decisions the *visit-shop-buy* (Sorensen 2016; Hui et al. 2009a). Using this double-conversion process we can divide the dynamical behavior of shoppers as in **Figure 5**.

We let n_k be the total number of visits of shopper k visits, v_{kt} be the node the customer k visits at time t and $v_k = \{v_{k1}, \dots, v_{kn_k}\}$ denote the shopping path of the customer k (visited nodes).

In this model, we distinguish the term the *zone* and the *node*. The *zone* is used to describe the physical segments of the floor as described in Figure \ref{floor}, whereas the *node* is used to describe a virtual sequence of shopper's path. We assume that each node locates only *one* item. For example, if shopper k visits the zone i to locate two items, say a and b , at time t , we define two (dummy) nodes v_{it} that locates the item a and $v_{i(t+1)}$ to locate the item b .

We let s_{kt} denote the dummy node denoting the customer k shops at time t and b_{kt} denote the dummy node denoting the customer k buys at time t . At each step t , shopper k visited v_{kt} choose to shop and to buy. This decision can be denoted as

$$X_{kt}^s = \begin{cases} 1 & \text{if the customer } k \text{ shops at } v_{kt} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{kt}^b = \begin{cases} 1 & \text{if the customer } k \text{ buys at } v_{kt} \\ 0 & \text{otherwise} \end{cases}$$

We let $p_{kt}^s = \mathbf{prob}\{X_{kt}^s = 1\}$ denote the probability of the customer k shops at v_{kt} , $p_{kt}^b = \mathbf{prob}\{X_{kt}^b = 1 | X_{kt}^s = 1\}$ denote the conditional probability of the customer k buys at

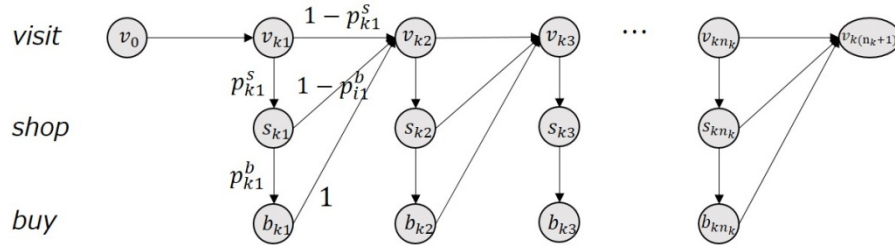


Figure 5 A Sequential Conversion Process

v_{kt} where the customer k shops at v_{kt} , and $p_{kt} = p_{kt}^s p_{kt}^b$ denote the double conversion rate. Although it is an important topic, we do not consider the heterogeneity of shoppers or items, as our interest lies in the average number of items purchased for all the shoppers in the store. Therefore, we take average probability of the conversion rate p , and assume that $p_{kt} = p, \forall k, \forall t$. Let $X_{kt} = X_{kt}^s X_{kt}^b = \{0,1\}$ be the random variable to take one if k -th shopper buys at zone v_{kt} and take zero otherwise, and x_{kt} be the realization of X_{kt} . Therefore, this process can be interpreted as the Bernoulli process with n_k times trial.

Let Y_{kt} be the number of items bought by the shopper k and n_b denote its realization. To estimate the parameter p , the likelihood of the Bernoulli process is defined as the joint probability of $[X_{k1}, \dots, X_{kn_k}]^T = [x_{k1}, \dots, x_{kn_k}]^T$ and the likelihood and the log-likelihood function of Bernoulli process is as follows:

$$\begin{aligned} L(X_{k1}, \dots, X_{kn_k} | p, n_v) &= \prod_{k=1}^{N_i} p^{x_k} (1-p)^{1-x_k} \\ &= p^{\sum_{k=1}^{N_j} x_k} (1-p)^{\sum_{k=1}^{N_j} (1-x_k)} \\ &= p^{n_b} (1-p)^{(n_v-n_b)} \end{aligned}$$

$\log L(X_{k1}, \dots, X_{kn_k} | p, n_v) = n_b \log(p) + (n_v - n_b)(1-p)$
With $d \log L / dp$, we have

$$\begin{aligned} n_v/p - (n_v - n_b)/(1-p) &= 0 \\ \Leftrightarrow p &= n_b/n_v \end{aligned}$$

Therefore, we can estimate the number of items purchased given the shopping-path as the binomial distribution as

$$\text{prob}\{y_k = n_b | n_v, p\} = n_v C_{n_b} (1-p)^{n_v-n_b}$$

The expected number and variance of n_b then are calculated as

$$\begin{aligned} \mathbb{E}(n_b | p, n_v) &= p n_v \\ \mathbb{V}(n_b | p, n_v) &= n_v p (1-p) \end{aligned}$$

The important observation is that not only the expected value $\mathbb{E}(n_b | p, n_v)$, but the variance $\mathbb{V}(n_b | p, n_v)$ depends on the shopping-path length n_v . Therefore, we can reason the fact that “the longer distance, the greater the variance”.

5. RESULTS

We see that the coefficient of n_v in both models, *i.e.*, $p = 0.0758$ and $\beta = 0.0744$, are coincidentally closed each other. This fact supports that our assumption of Bernoulli process modeling is appropriate.

Again, we see that not just the expected value, but the variance of n_b depends on the shopping-path length n_v , whereas in the normal regression assumption, the variance is always σ^2 .

With normal approximation $n_b \sim \mathcal{N}(n_v p, n_v p(1-p))$, we can have approximated upper bound (AUB) and lower bound (ALB) of γ -confidence region as

$$\begin{aligned} AUB &= n_v p + z \sqrt{n_v p(1-p)} \\ LUB &= n_v p - z \sqrt{n_v p(1-p)} \end{aligned}$$

where z is the quantile $\Phi^{-1}(\gamma)$ of the standard normal distribution.

We draw the 95% confidence region with $z = 1.96$ onto the scatter diagram as in Figure 6. We see that most of the points are within the bounds.

Finally, we discuss the managerial insight from this result. First, in order to increase the unplanned buying, it is important to know what factors that affect the number of purchases. The proposed Bernoulli process modeling includes the dynamical behavior of shoppers. The proposed approach divides a shopping path into a series of three exhaustive sequential and interrelated decisions the *visit-shop-buy* (Sorensen 2016; Hui et al. 2009a). Using this double-conversion process we can divide the number of items purchased as follows:

$$n_b = p_s \times p_b \times n_v$$

This distinction is important, because it allows us to take different in-store marketing activities depending on which of conversions to fail. The first measure n_v plays a critical role in the effectiveness of in-store marketing activities, while the other two measures p_s and p_b can be rather accounted for by products' attractiveness itself. To estimate n_v , it is necessary to understand shopping behaviors of customers deeply.

Second, the results justify the relationship that “the longer the distance, the more items purchased”. There is a strong consensus among both academia and industry that many purchases in a retail store are unplanned. Underhill (1999) claimed that “Supermarkets are places of high impulse buying..., 60 to 70 percent of purchases there were unplanned, grocery industry studies have shown us”. The results indicate the importance of exposure is very important for generating unplanned (impulse) purchases. To be influenced by in-store stimuli, the travel-distance plays a critical role. There are several shopper marketing strategies aimed at increasing travel distance such as placing popular products in scattered locations throughout the store

(Granbois 1968), or in-store mobile promotion strategies (Hui 2013).

Finally, from the result, we see that the need to be aware of the effect of the variance is examined. Systematic efforts to describe the characteristics of the variance in trip length have not been made despite its relevance to the number of item purchased. Variance presents a problem for retailers as it suggests that shoppers have different needs and meeting those needs with a single approach or layout may be difficult.

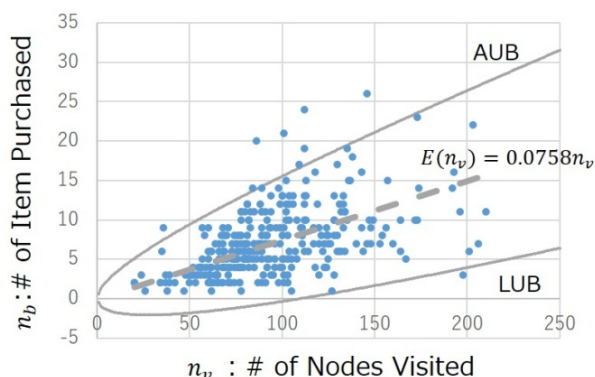


Figure 6 Scatter Diagram between # of Nodes Visited and # of Items Purchased

6. CONCLUSION

The length of shopping-path is a key metric for managing in-store merchandising as it allows shoppers to exposure to in-store stimuli to remind them of potential shopping needs. In this study, we analyzed the shopping-path data and found the fact "the longer the travel distance, the greater the variance of the number of items purchased". To explain this fact, we proposed a probability model of sequential purchase decisions using the Bernoulli process. We can estimate the number of items purchased as the binomial distribution given the shopping path.

For future research direction, we can analyze behavior of shoppers who bought many items even though the path-length is average. We should also consider the heterogeneity of shoppers or items. By segmenting them, we can get more accurate path-length estimation.

ACKNOWLEDGEMENT

We wish to thank the Distribution Economics Institute of Japan for providing the data set used in this paper.

REFERENCES

- Bell, D. R., Corsten, D., & Knox, G. (2011). From point of purchase to path to purchase: How preshopping factors drive unplanned buying. *Journal of Marketing*, 75(1), pp. 31-45.
- Granbois, D. H. (1968). Improving the study of customer in-store behavior. *The Journal of Marketing*, pp. 28-33.
- Hansen, K., & Olsen, S. O. (2006). Impulsive Buying Behavior-The Role of Impulse Buying Tendencies and Convenience Orientation. In ANZMAC 2007 Conference, University of Otago, New Zealand, December 3 (Vol. 5).
- Hui, S. K., Bradlow, E. T., & Fader, P. S. (2009a). Testing behavioral hypotheses using an integrated model of grocery store shopping path and purchase behavior. *Journal of Consumer Research*, 36(3), pp. 478-493.
- Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009b). Path data in marketing: An integrative framework and prospectus for model building. *Marketing Science*, 28(2), pp. 320-335.
- Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009c). Research note—the traveling salesman goes shopping: The systematic deviations of grocery paths from TSP optimality. *Marketing Science*, 28(3), pp. 566-572.
- Hui, S. K., Inman, J. J., Huang, Y., & Suher, J. (2013). The effect of in-store travel distance on unplanned spending: Applications to mobile promotion strategies. *Journal of Marketing*, 77(2), pp. 1-16.
- Inman, J. J., Winer, R. S., & Ferraro, R. (2009). The interplay among category characteristics, customer characteristics, and customer activities on in-store decision making. *Journal of Marketing*, 73(5), pp. 19-29.
- Kahn, B. E., & Schmittlein, D. C. (1989). Shopping trip behavior: An empirical investigation. *Marketing Letters*, 1(1), pp. 55-69.
- Knox, G., Bell, D. R., & Corsten, D. (2011). Situational determinants of unplanned buying in emerging and developed markets. Marketing Science Institute Working Paper Series. Marketing Science Institute.
- Kollat, D. T., & Willett, R. P. (1967). Customer impulse purchasing behavior. *Journal of Marketing Research* 4(1), pp. 21-31.
- Rook, D. W. (1987). The buying impulse. *Journal of Consumer Research*, 14(2), pp. 189-199.
- Underhill, P. (1999). *Why we buy: The science of shopping—updated and revised for the Internet, the global consumer, and beyond*. Simon and Schuster.
- Sorensen, H., (2009). *Tell 'em Where to Go; Tell 'em Which to Buy!*, Views. TNS Global, Online.
- Sorensen, H. (2016). *Inside the mind of the shopper: The science of retailing*. FT Press.
- Sorensen, H., Bogomolova, S., Anderson, K., Trinh, G., Sharp, A., Kennedy, R., & Wright, M. (2017). Fundamental patterns of in-store shop-per behavior. *Journal of Retailing and Consumer Services*, 37, pp. 182-194.
- Stilley, K. M., Inman, J. J., & Wakefield, K. L. (2010). Spending on the fly: mental budgets, promotions, and spending behavior. *Journal of Marketing*, 74(3), pp. 34-47.
- Tanskanen, K., Yrjola, H., & Holmstrom, J. (2002). The way to profitable Internet grocery retailing—six lessons learned. *International Journal of Retail and Distribution Management*, 30(4), pp. 169-178.
- Walters, R. G., & Jamil, M. (2003). Exploring the relationships between shopping trip type, purchases of products on promotion, and shopping basket profit. *Journal of Business Research*, 56(1), pp. 17-29.
- Zhang, X., Li, S., Burke, R. R., & Leykin, A. (2014). An examination of social influence on shopper behavior using video tracking data. *Journal of Marketing*, 78(5), pp. 24-41.

Shunichi Ohmori (Ph.D) is an assistant professor at department of industrial & system engineering, Waseda University in Japan, and a researcher at institute of global production & logistics at Waseda University, and a researcher at data science institute at Waseda University. He received the master and Ph.D degree in engineering at Waseda University. His research interest lies in operations research and supply chain management.

Masao Ueda (Ph.D) is a professor at department of business design management, Waseda University in Japan, and a researcher at data science institute at Waseda University. He received the master degree in agriculture at Hokkaido University and Ph.D degree in commerce at Waseda University. His research interest lies in marketing engineering and marketing research.

Kazuho Yoshimoto (Dr. Eng) is a professor at department of industrial & system engineering at Waseda University in Japan, and a head of Institute of Global Production & Logistics at Waseda University. He received the master and Ph.D degree in engineering at Waseda University. His research interest lies in facility and logistics design.