RIETI Discussion Paper Series 17-E-035

On the Role of Skill, Quality, and Environmental Factors on Customer Behavior of the Beauty Industry

KONISHI Yoko
RIETI

The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Paper Series 17-E-035

March 2017

On the Role of Skill, Quality, and Environmental Factors on Customer Behavior of the Beauty Industry*

KONISHI Yoko

Research Institute of Economy, Trade and Industry

Abstract

What do hair salons and hairdressers provide for their customers? One answer is services and creation of customer satisfaction. Customer satisfaction could depend on factors such as location, price, the skill of the hairdressers, and the overall experience at the hair salon. It is difficult to observe customer satisfaction directly for hair salon owners and researchers, because satisfaction is subjective. However, if the customers like the hair salon or hairdressers and are satisfied with their services, customers would return at a higher frequency and become big spenders. This suggests that we can measure customer satisfaction by investigating what causes them to return. In this study, we apply count process and double hurdle analysis to specify a model of customers' return and purchase behavior. We collected daily records from a hair salon in Japan between 2003 to 2010 (2,046 working days). The hair salon has about 15,000 customers and the daily records contain payments and the hairdresser's names for each customer visit. Using the information, we introduced customer behavioral variables, the hairdresser's skill and how crowded the salon is in addition to demographic variables into our model.

Keywords: Personal service industry, Hair salon, Double hurdle model

JEL classification: D57, F14, F68, L83

*This study is conducted as a part of the "Decomposition of Economic Fluctuations for Supply and Demand Shocks: Service industries" project undertaken at RIETI. The authors are grateful for helpful comments and suggestions by Senior Fellow Yoichi Sekizawa (RIETI), Professor Jiro Nemoto (Nagoya University), Professor Tadahiko Sato (Tsukuba University), Professor Yoshihiko Nishiyama (Kyoto University), Professor Naohito Abe (Hitotsubashi University) and participants of "Statistical Modeling for Marketing Seminar" at Tsukuba University and Hitotsubashi-RIETI International Workshop on Real Estate Market, Productivity, and Prices. Also, the authors appreciate invaluable advice and comments by Professor Kyoji Fukao (Hitotsubashi University, RIETI-PD), Vice President Masayuki Morikawa (RIETI), President and CRO Makoto Yano (RIETI) and seminar participants at RIETI. This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Grant-in-Aid for Scientic Research (No. 15K03407) and Research Project 2016 of Kyoto Institute of Economic Research.

email: konishi-yoko@rieti.go.jp

1. Introduction

According to the Cabinet Office's National Accounts of Japan, looking at the composition of the Japanese economy in recent years, the reality is that non-manufacturing industries account for 75% or more of GDP. More specifically, the service-providing industry's share of GDP stands around 20% and, when the transport, wholesale, and retail trade industries are included, this figure rises to 35% or more, consistently exceeding the manufacturing industry's share since 2009. In this paper, the term "service-providing industry" collectively refers to the narrowly-defined service industries classified into categories L through R under the Japan Standard Industrial Classification (13th revision in October 2013). It includes the personal, business and public services such as lease and rental business, restaurant, hair salon and barber shop, hotel, laundry, education, medication, professional expertise and so on. Also, the service-providing industry's share of the labor market has been on the rise since the 1990s and reached approximately 30% by 2009, whereas the manufacturing industry's share has trended downward since the 1990s to approximately 15% in 2014.

As evident from these figures, overall economic activity can no longer be explained simply in terms of the manufacturing industry. The service sector is extremely important due to its presence in the overall economy and magnitude.

According to "the Economic Census for Business Activity" conducted by Ministry of Economy, Trade and Industry (METI) of 2012, the hair salons was the first place about the number of establishment and the number was 170,000 stores. At the same time, the barbershop was the fourth place. Surprisingly, the establishments that specialize in cutting and care the hair is more than 270,000 shops in Japan. The second place was the drinking places (220,000 shops) and the third place was real estate (176,000 offices). Restaurants and convenient stores often seen around our daily life were the ninth place (55,000 shops) and 25th place (30,000 stores), respectively.

Here, special attention should be given to how we define "productivity" in the service sector. A typical measure of productivity--which has been used in many cross-country comparisons, government reports, and research works--is labor productivity, i.e., sales or value added per worker. Another measure of productivity, which takes into account the effects of input factors other than labor, is total factor productivity (TFP). Input factors refer to all factors involved in production such as the number of workers, physical capital (machinery, equipment, and facilities), raw materials, and energy. It is also more desirable to use TFP rather than labor productivity for the service sector. However, as aforementioned, economics has long failed to recognize the significance of this sector, and the smaller size of service-sector establishments--as compared to those in the manufacturing industry--makes it difficult to collect detailed data (particularly on machinery, equipment, and facilities). Under these circumstance, Morikawa (2010) provides a very valuable research that examine both quantity (physical) TFP and monetary (revenue) based TFPs for Japan's personal service industry by using establishment level data. In order to remove the demand effects from examined TFPs, he utilize some demand effects variables (e.g. market size, the degree of competition in the market and the day of week information) in the production function estimation.

As mentioned above, the service industry is a very broad classification, which includes education, finance, insurance, transportation, logistics, food service and many other sectors. These exhibit, in our understanding, distinctive different structures, and the construction of a single model accounting for all service sectors does not appear to be an easy feat. When TFP or labor productivity is not sufficient as a measure of productivity for some industries, we must identify and define how and what services and goods are provided in each industry and what sort of value added such services and goods generate. After that, and only then, we should collect data on both the demand and supply sides for empirical analysis. Let's take a look at the hair dressing and beauty salon industry. In Konishi and Nishiyama (2010), the value added in hair salon is defined as an improvement in the appearance of customers after hairdressing services. When haircutting and other technical skills are improved, time required for such services will be reduced, resulting in an increase in the quality of services per unit hour, greater customer satisfaction, and a greater probability for repeat customers. It also translates into an increase in the productivity of those hairdressers and hence an increase in the salon's sales. In this research, they used a demand- and supply-side model capable of simultaneously explaining customers' behavior and a production function that incorporates standard time required to perform a haircut as an input. Konishi (2010) apply duration analysis to specify a model of a hair salon customers' behavior regarding their visits and then estimate their aggregated revisit rates/dates. They estimate the intensity function by adopting the Cox model for hair salon data to examine which demographic, geographic, behavioral variables and supply side's information influence each customer's attitude.

According to Konishi (2010) and Konishi and Nishiyama (2010), we could obtain an important perception such that hair salon industry provides services and create customer

satisfaction. The customer satisfaction could depend on factors such as location, price, the skill of the hairdresser, and the overall experience at the hair salon. It is difficult to observe customer satisfaction directly for hair salon owners and researchers, because satisfaction is subjective. However, if the customers like the hair salon or hairdressers and are satisfied with their services, customers would come back to the hair salon and come back more often. This suggests that we can measure customer satisfaction by investigating customer's behavior.
In this study, our main research concern is the customer's revisit behavior and we have four research objects as below.

1. identify the difference at the first visit between a chance customer and loyal customer
2. find impact factors of demand (satisfaction) function
3. the reasons of loyalty or life time value gain
4. measuring the effects of skill of hair dressers and the environment of hair salon on customer satisfactions

To investigate these objects, we need to count each customer's number of visit, secondly, we need to examine the amount of customer's repurchase and thirdly, we need to investigate the customer's lifetime value. These three variables are outcome and we adopt the count process analysis and double hurdle analysis as below. Because more than 40% customers didn't come back to the hair salon (See Figure 3.), the dataset has excess zeros (non-revisit and no-purchase) observation. Usually, excess zeros caused overdispersion and we should find alternative estimation methods other than the Poisson regression for the number of revisit data. As alternative ways, there exist the zero-inflated Poisson (hereafter, ZIP) model, the zero-inflated negative binomial (hereafter, ZINB) model. In our empirical analysis, we will apply ZIP and ZINB regression to revisit events. For repurchase data, most famous censored dependent variable model is a Tobit model, it assumes that the decision of revisit the hair salon and how much purchase in the next visit are explained by same factors. On the other hand, double hurdle model proposed by Cragg (1971) allows two decisions are different. The double hurdle can be expressed by combination of probit and tobit estimator for decision of revisit and amount of purchasing, respectively.
We collected a daily record from a hair salon in Japan between 2003 to 2010 (2,046 working days). The hair salon has about 15,000 customers and the daily records hold payments and hairdresser's names for every customer's treatment. In order to reproduce the customer

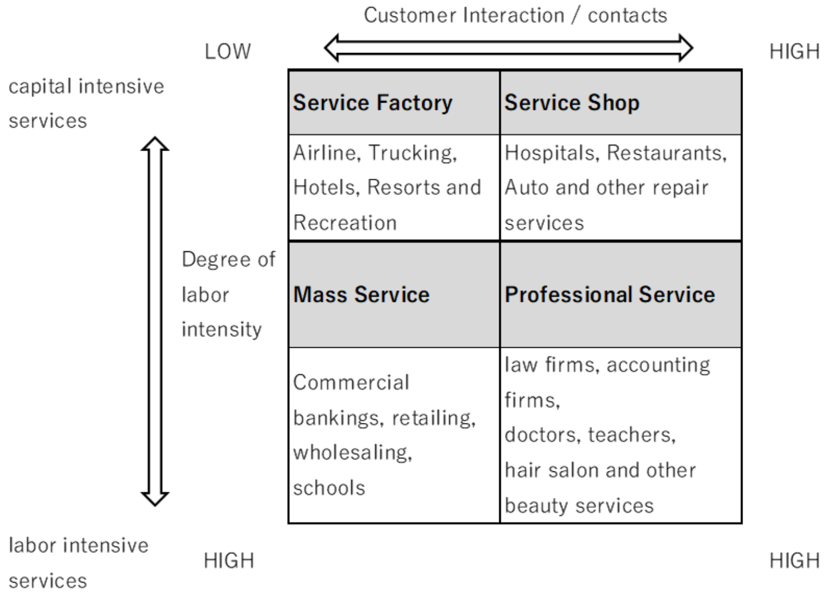
experience space at hair salon, we introduce customer behavioral variables, the hairdresser's skill and congestion (supply side data), and the hair salon's congestion (hair salon's atmosphere or environment) in addition to demographic variables into our model.

The following section presents some related literatures about concept of customer satisfaction and service quality, briefly. Section 3 shows preliminary analysis regarding the summary of the variables through an empirical study. Section 4 introduces the estimation methods for count process and quantity data. In Section 5, we provide the empirical model and explanations of covariates, then discuss about the estimation results. The concluding remarks are given in Section6.

2. Customer Satisfaction and Service Quality

In this section, we introduce some related literatures about concept or definition of customer satisfaction and service quality in the marketing, management and psychological field, briefly. For many years, numerous marketing researchers have attempted to classify the service industry and the quality of services in order to know what the dominant factors and measure service quality. Schmenner (1986) breaks service industry down into four patterns by using the degree of labor intensity of the customer interaction of the service industry. The labor intensity denotes that the ratio of labor cost against value-added. The customer interaction means that the frequency of service providers' contact to customers during the services. Additionally, it includes the generous and flexible actions of service providers to customers' requests. Hairdresser requires the qualification and they keep interacting with their customers during the treatments. Thus, we think they should be allocated on professional services in Figure 1.

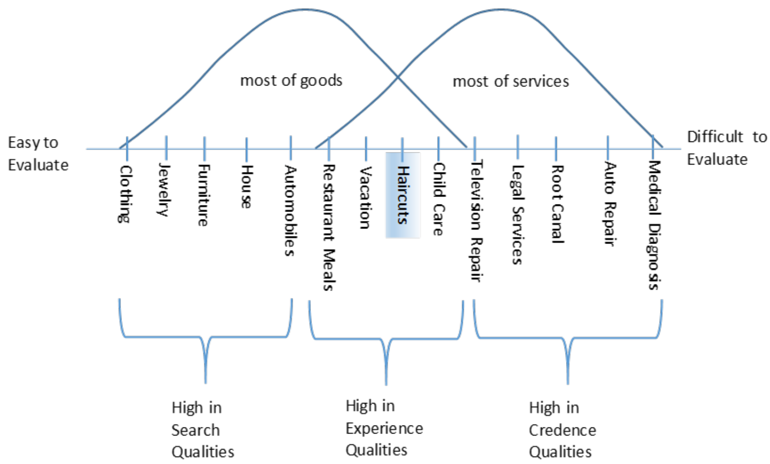
Figure 1: Classification of Services



Source: Schmenner (1986) and add some modification by author

Nelson (1971) provided a pioneering work about the evaluation for the quality of consumer goods by introducing two attributes of qualities such as "search qualities" and "experience qualities". Kani (1973) adds one attribute on Nelson's idea, that is "credence qualities". Furthermore, Zeithaml (1981) indicates the evaluation processes of qualities are differ between consumer goods and services by using the three attributes (see figure 2). Goods and Services are allocated high in search, experience and credence qualities along a continuum of evaluation ranging from easy evaluation to difficult evaluation. Firstly, the easiest evaluate qualities of goods is linking with the attribute of "high in search qualities", because the goods can be evaluated the quality even before purchase (e.g. clothing, foods, Jewelry). Secondly, "high in experience qualities" for both goods and services are more difficult to evaluate, because they must be purchased and consume before assessment (restaurant meal, haircuts, childcare). Finally, some services are belonged to "high in credence qualities", they are most difficult to evaluate, because the consumer may be unaware of or may lack sufficient knowledge to appraise whether offering satisfy given wants or needs after consumption (e.g. legal services, medical diagnosis). In their works, haircuts are regarded as "high in experience qualities". Also, Zeithaml (1981) pointed out that it allows to vary with time if consumer accumulate their knowledge and experience of the goods or services after purchasing.

Figure 2: Continuum of Evaluation for Different Types of Products



Source: Zeithaml (1981), p.186

Grönroos (1988) are defined the quality of services as the technical aspect ("what" service is provided) and the functional aspect ("how" the service is provided). Also Grönroos indicates the relation between the two aspect and six components of service quality. "Technical quality" represents "professionalism and skill", and "Functional quality" relates to "attitudes and behavior", "accessibility and flexibility," "reliability and trustworthiness," "recovery and reputation". The sixth component is "reputation and credibility" which link with the quality of image for service provider.

As other representative works in this field, Parasuraman, Zeithman and Berry (1985) developed SEVQUAL, it consists of 10 components to evaluate service quality. Afterwards, many researchers have utilized SEVQUAL as a means to examine the quality of services. Also, SEVPREF proposed by Cronin and Taylor (1992), which measures service quality by using customer's evaluation for the "actual provided" services other than expected services.

In the service industry, the service delivery environment other than technical skill has been pointed out to be related to customer satisfaction (Hussain and Ali (2015)) . Usually, in Japan, we could say that state of the technical skills of the hairdresser is sufficiently high. Thus, in order to further increase customer satisfaction, there will be necessary to increase the functional quality proposed by Grönroos (1988).

Generally speaking, service industries possess the following characteristics.

1) Simultaneity: Services are consumed as provided.

2) Inseparability: There is no partial receipt of the service; there is no separating the place of consumption from the place of provision.

3) Intangibility: No inventories are held; inventories are invisible.

Furthermore, we would like to add one feature for personal services.

4) Heterogeneity: Personal services such as treatment directly to the hair and the body, even if the same provider to provide the same services that are the outcome quality must be different.

As discussion above, marketing and management researchers have recognized hair salon (personal services) as conducive business to study. On the other hand, in economics research, hair salon and barber shop are cited as example of industry which have very simple structure. For example, hair salon and barber shop are mostly domestic players and tend to be "locally oriented" in their operations, because of these 1)-3) characteristics. Additionally, because hair salon and barber shop have less product differentiation, thus consumers decided a shop by whether the shop is proximity or not (Ellickson (2007), Kolko (2007) and Oi (1992)).

However, focusing on the customer's experiences at the hair salon, the customer must tolerate the closeness of the distance between the hairdressers, that there are the scissors and razor near the face, the line of sight of the other customers, also the conversation with the first meeting or not-so-friendly people, thus there are many to be a stress factor. If the customers feel fear, stress and dissatisfaction during the process of services, they must evaluate the functional quality of hair salon is low. If the hair salon understands the value and preferences of functional quality for each customer and then they can provide comfortable environment, it will lead to raise of customer loyalty and obtain the service differentiation.

Oliver (1993) described that customer satisfaction is a mixed of cognitive and affective components. Also, in Oliver (1999), the customer satisfaction is defined as "the consumer's fulfillment response, the degree to which the level of fulfillment is pleasant or unpleasant". Salon's owner, hairdressers and researchers cannot observe precisely such satisfaction or dissatisfaction and pleasant or unpleasant. However, there surely exist. Recently, the concept of "Switching Barrier" (Jones, Mothersbaugh and Beatty (2003) and many others) helps to understand the relationship between unobservable customer's mind and their behavior by searching a cause of behavioral change. For example, customers remain in the hair salon even if there is dissatisfaction, but the customer gets away the hair salon when dissatisfaction is larger than the cost of switching to the other hair salon. Also, in the recent years of psychological research (Gross (2015)), a concept called "Emotion Regulation" is extensively discussed. This refers to the action to prevent the occurrence of unpleasant feelings, it is also included the customer's decision such as to avoid the hair salon they do not want to go. In our model, we would like to focus on the role of functional quality as well as technical quality to observe the factors of hair salon's demand decisions.

3. Data and Preliminary Analysis

3.1 Descriptions of a Hair Salon

We obtained data regarding customer attributes and customer purchasing activities from a hair salon. The hair salon is located in Osaka city in Japan. The salon is easily accessible from multiple railway stations, business offices, restaurants, and shopping malls. Nearly all Japanese hair salons require the customer to fill in a questionnaire on their first visit. Salon owners can obtain the data on the customer's gender, age, occupation, address, hobby and some other preferences. In order to retain the customers, salon owners send direct mails with coupons at regular intervals or customer's birthday. Meanwhile, knowing the customer information from the questionnaire helps to capture the latent preferences of the customer, also it helps to take a smooth communication with customers in the conversations.

We collect an unbalanced panel data of individual ＊ item (hair goods and services) ＊ day.

The hair salon opened in July 2003, and we have observed them from July 2003 to March 2010 (2,046 working days). The hair salon provides the customers with various hair products (shampoo, conditioner, hair oil, Hair spray and so on) and services including haircuts, color, permanent waves, special care, hair set and arrangement and more. As you know, their main services are haircuts, color, and permanent waves and their share of total number of treatments are 87%, also the three services yield 90% of the total sales. The hair salon has sold around 90.0milion yen annually from 2004 to 2009.

In this paper, we use the customers' gender, age and the distance between the salon and their home addresses as demographic variables from questionnaires and behavioral variables from daily customer purchasing activities. They have 15,942 customers and female customers account for 90.5%. Table 1 shows a change of number of customers from 2003 to 2010 and summary statistics of age of customers after removing the customer who are more than 100-year-old or less than 10-year-old. The averages age of customers are around 26 years old and the 75 percentiles are 30-year-old during the period, thus, majority of customers lie between 20 and 30 years of age. The distance between the residences of the customers and the salon can be very skewed to the right (See Table 2). The mean and median of distance is 14.6km and 7.3km, respectively. It is surprised that the maximum distance is more than 1,000 km, and 203 person of customers come through the hair salon away from more than 100km. We observed forty prefectures out of forty-seven prefectures in their addresses. This might be due to the fact that the customers happened to visit the salon while they had some other tasks planned in the neighborhood, or because they wrote down another address, e.g., parent's address since they dislike receiving many direct mails. In any case, it is difficult to regard these customers as frequent customers and hence we treat them as outliers. After we abstracted the customers who live within 100km or 50km from the hair salon, about 1.6% or 2.2% of the customers are removed. The remaining about 98% customers are from four prefectures (Osaka, Kyoto, Hyogo and Nara) and 88% of the customers are from Osaka prefecture.

As we mentioned, we can use these demographic variables (gender, age, distance btw. Residence and the salon) from the initial questionnaire. We have another information about the daily records of customers' visits. The daily records hold the payment and hairdressers' name for each customer's treatment, making it possible to aggregate the experiences of hairdressers for the seven years and examine both hairdressers and the salon's capacity, skill, productivity and/ or quality of services. In order to specify the model of customers' satisfaction, we use not only the customers' demographic variables but also customers' behavioral variables and the supply side (hairdressers and salon) variables which are made by using daily records.

Table 1. Number of Customers and Summary Statistics of Customer's Age (2003-2010)

Year	Number of Customer	Mean	Median	Std. Dev.	Min	Max
2003	2,114	26.4	25	7.7	11	77
2004	3,100	26.3	25	7.8	12	75
2005	2,558	26.7	25	7.4	11	75
2006	1,767	27.0	26	7.2	11	75
2007	1,951	27.4	26	7.4	10	77
2008	1,877	27.1	26	7.1	13	75
2009	1,156	27.2	26	7.6	10	74
2010	348	25.9	25	6.9	12	65
Total	14,871	---	---	---	---	---

Table 2. The Distance between the Residences of the Customers and the Salon (km)

# of Customer	Mean	Std. Dev.	Median	75percentiles	95percentiles	min.	max.
12,799	14.6	48.0	7.3	12.5	33.1	0.98	1264

3.2. Data Construction and Some Key Variables

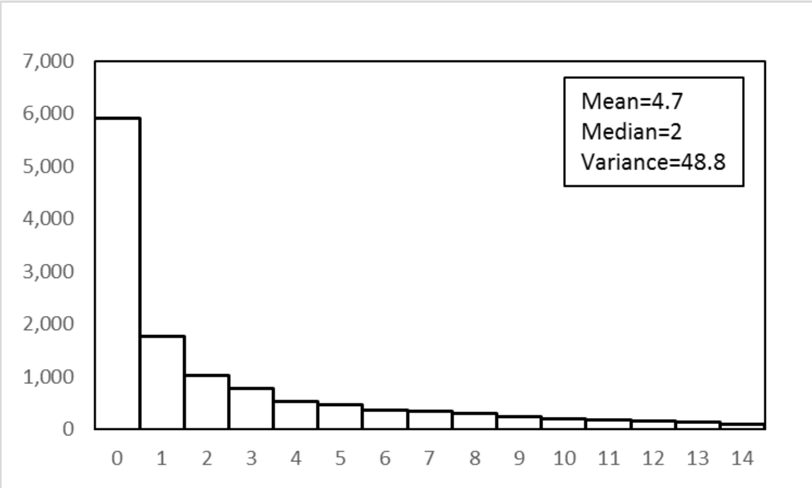
In order to construct a target dataset for our analysis, first, we abstract from the full dataset, the data on the customers who used three (haircut, hair coloring and permanent wave) services and then remove the customers with the following conditions: [1] customers who live 100km away from the hair salon; [2] customers who cannot be identified as individuals, e.g. id number, address and birthday are missing After abstracting our target dataset, 12,596 customers remain (full data set has 15,942 customers).

We have three research objects as below.

1. identify the difference at the first visit between a chance customer and loyal customer
2. find impact factors of demand (satisfaction) function
3. the reasons of loyalty or life time value gain
4. measuring the effects of skill of hair dressers and the environment of hair salon on customer satisfactions

To investigate these objects, we need to count each customer's number of visit, secondly, we need to examine the amount of customer's repurchase and thirdly, we need to investigate the customer's lifetime value.

Figure 3. Histogram Number of Revisits (2003-2010)



* Notes: we truncate the sample at 15 revisit times.

Figure 3. is a histogram of number of revisit times, we can see excess zeros and it skewed to right. It seems to be like Poisson distribution, but this is overdispersed (mean<variance). Number of non-revisit customers account for 44.3% of sum of customers. Number of revisit customers will be outcome variable in Section 5.2.

Table 3. Number of Customers and the Average Customer Spend by segments

Customer Segmentation	total visit times	# of customers (Share)	average customer spend
non revisit customer	1 time	5903 (44.3%)	¥7,710
regular customer	2-4 times	3560 (26.7%)	¥8,068
loyal customer	5-10 times	2216 (16.6%)	¥8,924
super loyal customer	11 and more	1645 (12.3%)	¥9,460
Total		13,324	¥8,223

We classified customers into four segmentations by the total visit times in Table 4. The customers who visited at the hair salon 4 or less accounts 71%. The average customer spend of super loyal customer is more than 120% of non-revisit customer's one. As the number of total visits increases, the average customer spends are getting higher. We might say the loyalty is growing.

Table 5. Duration and Average Customer Spend by kind of services

duration of non revisit (day)	Obs.	mean	median	std.dev.
Hair Cut	15,531	86.0	61	107.9
Hair Coloring & Cut	17,686	86.8	69	85.7
Parmanet Wave & Cut	7,304	89.4	64	99.8
average customer spend (yen)	Obs.	mean	median	std.dev.
Hair Cut	19,096	4380.6	3780	1769.4
Hair Coloring & Cut	22,540	11281.1	10920	2973.7
Parmanet Wave & Cut	10,101	11585.6	10950	3626.3

Table 5 shows that the summary statics for the duration of non-revisit and the average customer spend by kind of services. Among the three combination of services, we can see they have similar intervals of revisiting. On the other hand, the average customer spends look different, the average customer spends of hair cut is only about one-third of other two combinations. It is very important to raise the amount of sales by recommending an additional service to the customers.

3.3 Measurement Skill, Productivity and Capacity

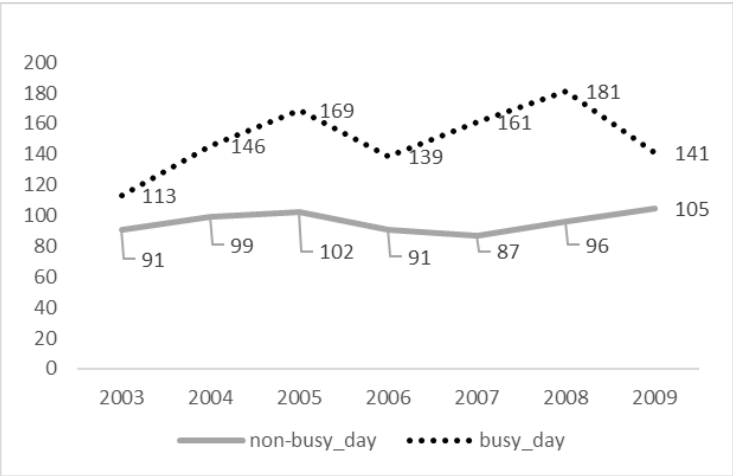
Konishi and Nishiyama (2010) define one of the productivity growth that "the expansion of capacity of hairdresser and/or hair salon." They showed that the nonparametric identification and estimation of capacity. They discussed that each hairdresser has his/her own capacity on how many haircuts they can give in one day. Number of haircuts in a day is determined as a minimum of the capacity and the demand of the day. In view of the data, however, it may be more suitable to assume that the capacity may change day by day, say, depending on the hairdressers' conditions or other external reasons. Also, they may sometimes need to operate below the predetermined level of their own quality in order to meet the too many demands, though they basically do not do it. Thus we rather assume it is not completely fixed. In this hair salon, they sometimes reject demands from customers on busy days such as Fridays, Saturdays, Sundays, national holidays, and other special days as around Christmas and new year days. But it never happens on non-busy days such as ordinary Tuesdays, Wednesdays and Thursdays. This may help us to identify the capacity as a set. We can think the capacity is greater than the maximum of customer number on non-busy days. Also, the standard (or average) capacity must be smaller than the maximum number in a whole year. Therefore, we can estimate the capacity by the set;

$$\left(\max_{d \in non\_B_y} n_{id}, \max_{d \in B_y} n_{id} \right) \cdots (1)$$

$$\left(\max_{d \in non\_B_y} n_{d}, \max_{d \in B_y} n_{d} \right) \cdots (2)$$

Here B_y is the set of busy days in year y and non_B_y is that of non-busy days, n_{id} denotes the number of customers for hairdresser i. Eq. (2) is for hair salon's capacity equation. We need to be careful in interpreting the upper bound as it depends on the model of random capacity at least in part, but we believe the lower bound is very reliable. We showed the results for aggregated hair salon in Figure 4. The set looks to move upward with the accumulation of experiences. This indicates the possibility that the technical skill of hair salon increases with experience and the capacity increases.

Figure 4. Estimate Results of Hair Salon's Capacity



In order to observe productivity and technical capabilities in hair salon (the service sector), it is important to obtain daily data on the maximum amount of services and standard time spent for performing a haircut. However, what we can usually observe is the lesser of the maximum amount of services available and actual demand for services. For instance, a hairdresser capable of providing 20 haircuts during the opening hours would not spend two hours per customer even when there are only four customers. However, the fact that there were only four customers is all that is visible to analysts. We would be obviously underestimating the hairdresser's technical capabilities if we interpret this observation as an indication of his or her being able to provide only four haircuts per day. To solve this problem partially, we examine the both hair salon and hair dresser's capacity and use them as measures of skill and capacity variables in our empirical analysis in Section 5.

4. Estimation Methods

4.1 Models for Count Data: Poisson model and Zero Inflation Models

Our main research concern is the customer's revisit behavior, and events of revisit are count variables. As shown Figure 3. in previous section, because more than 40% customers didn't come back to the hair salon, the dataset has excess zeros (non-revisit and no-purchase) observation. Usually, excess zeros caused overdispersion and we should find alternative estimation methods other than the Poisson regression. As alternative ways, there exist the zero-inflated Poisson (hereafter, ZIP) model, the zero-inflated negative binomial (hereafter, ZINB) model and the (double) hurdle model. In our empirical analysis, we will apply ZIP and ZINB regression to revisit events and apply the hurdle model for quantity (monetary) data of the amount of customer's purchases. In order to describe about ZIP and ZINB models, we show the probability density function (PDF) of the Poisson and negative binomial distribution as below,

$$\mathrm{PDF}_{Poisson}(y;\lambda) = \frac{\exp(-\lambda_i)\lambda_i{}^{y_i}}{y_i!}$$

$$\mathrm{PDF}_{NB}(y;p,r) = \frac{(y_i! + r - 1)!}{y_i!\,(r-1)!} p_i^r (1-p_i)^{y_i}$$

, where y_i is outcome variable and generated from each count process, λ is parameter of the Poisson distribution, and p is probability of r success of negative binominal model.

When number of n observations are each independently, we observe the likelihood function of the Poisson distribution is given by

$$L(y;\lambda) = \prod_{i=1}^{N} \frac{\lambda^{y_i}}{y_i!} e^{-n\lambda},$$
$$y_i = 0, 1, \dots, i = 1, \dots, N$$

, also log-likelihood function is denoted as below.

$$\log L = \sum_{i=1}^{N} [y_i \log\lambda_i - \lambda_i - \log(y_i!)].$$

Meanwhile, negative binomial model's likelihood function and log-likelihood functions are taken by

$$L(\lambda; y, \alpha) = \prod_{i=1}^{N} exp\left\{y_i \log\left(\frac{\alpha\lambda_i}{1+\alpha\lambda_i}\right) - \frac{1}{\alpha}\log(1+\alpha\lambda_i) + \log\Gamma\left(y_i + \frac{1}{\alpha}\right) - \log\Gamma(y_i+1) - \log\Gamma\left(\frac{1}{\alpha}\right)\right\}$$

$$\log L(\lambda; y, \alpha) = \sum_{i=1}^{N} y_i \log\left(\frac{\alpha\lambda_i}{1+\alpha\lambda_i}\right) - \frac{1}{\alpha}\log(1+\alpha\lambda_i) + \log\Gamma\left(y_i + \frac{1}{\alpha}\right) - \log\Gamma(y_i+1) - \log\Gamma\left(\frac{1}{\alpha}\right)$$

, where α allows for overdispersion and we can link λ_i with $\exp(x_i'\beta)$ in both Poisson and negative binomial regression. α and β are unknown parameters, and are obtained by maximizing each log-likelihood function.

Many previous researches stated that modeling consumption of particular goods with expenditure household data is difficult due to excess zero observation in the sample. Usually, we might observe zeros come through three reasons such as; (1) customers decide not to buy, (2) customers had intention to buy, but they didn't purchase the goods or anything else (e.g. buy another good, just change mind and out of stock, etc.) and (3) when the goods are durable, the observing period is too short to obtain the purchase records. In hair salon or restaurant case, once customers enter the shop, they will buy the services or product certainly. In other words, zeros observations are originated by non-revisit, we don't need to consider the customers who visit the store but not buy. We deal only the case of (1) and (3).

After latest haircut, the intensity of haircutting is getting stronger over the time. If the customer decided to go back to the same hair salon, and then she / he decides on the quantity or amount of that services at the hair salon. Also, she /he regardless of whether change of mind or force majeure, they are allowed to change the revisit decisions. In ZIP and ZNIB framework, outcomes are divided into the two processes due to a customer revisit or not revisit at the hair salon. When a customer never revisits, y_i should be zero. On the other hand, when a customer revisits the hair salon, y_i follows count process. ZIP model has two parts that are logit model if outcome is zero and are Poisson model if outcome is larger than one. The log-likelihood function is given by

$$\log\mathrm{L}(\beta,\gamma) = \sum_{y_i=0} \log\left[\exp(z_i'\gamma) + \exp\left(-\exp(x_i'\beta)\right)\right] + \left[\sum_{y_i\neq 0} y_i x_i'\beta - \exp(x_i'\beta) - \log(y_i!)\right] - \sum_{i=1}^{n} \log[1+\exp(z_i'\gamma)] \cdots (3)$$

,where z_i and γ are exogenous variables and unknown parameter, respectively.

ZNIB model also has two parts, we implement logit regression for predicting zeros. When y_i is larger than one, we use a negative binomial count model. The log-likelihood function is such as

$$\log\mathrm{L} = \sum_{i=1}^{n} \begin{cases} \log(p_i) + (1-p_i)\left(\dfrac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}} if \;\; y_i = 0 & \cdots (4) \\ \log(p_i) + \log\Gamma\left(\dfrac{1}{\alpha} + y_i\right) - \log\Gamma\left(\dfrac{1}{\alpha}\right) + \left(\dfrac{1}{\alpha}\right)\log\left(\dfrac{1}{1+\alpha\mu_i}\right) + y_i\log\left(1 - \dfrac{1}{1+\alpha\mu_i}\right) & if \;\; y_i > 0 \end{cases}$$

, where $\mathrm{p} = \frac{1}{1+e^{-x_i'\beta}}$, $(1-\mathrm{p}) = \frac{1}{1+e^{x_i'\beta}}$.

ZNIB regression is more appropriate than ZIP regression, if the outcome is overdispersed. Which model is preferable or not should be an empirical issue, we apply both model for our revisit counts data and then implementing specification diagnostic by using estimation results.

4.2 Models for Quantity Data: Double Hurdle Model

As mentioned in previous sub-section, our main research concern is the customer's revisit behavior. In our sample, we have excess zeros (non-revisit and no-purchase) observation, that is, both revisit and repurchase outcomes are bounded. Most famous censored dependent variable model is a Tobit model, it assumes that the decision of revisit the hair salon and how much purchase in the next visit are explained by same factors. On the other hand, double hurdle model proposed by Cragg (1971) allows two decisions are different. Cragg (1971) provided that separate participation and quantity equations for bounded and the unbounded outcomes. The hurdle model concerns that the unbounded outcomes are over jump a hurdle. The double hurdle can be expressed by combination of probit and tobit estimator such as

$$d_i^* = z_i'\gamma + \varepsilon_{1,i} \ \ , y_i^{**} = x_i'\beta + \varepsilon_{2,i} \cdots (5)$$

$$\begin{pmatrix} \varepsilon_{1,i} \\ \varepsilon_{2,i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right]$$

, where $z_i\gamma + \varepsilon_{1,i}$ and $x_i\beta + \varepsilon_{2,i}$ represent the perticipation equation and the quantity equation, respectively. It assumes that the error terms of two equations are uncorrelated. The first hurdle is defined by the latent variable d_i^*,

$$d_i = 1 \ \ if \ \ d_i^* > 0$$
$$d_i = 0 \ \ if \ \ d_i^* \leq 0.$$

The second hurdle is given by

$$y_i^* = max(\ y_i^{**}, 0),$$

consequently, outcome variable y_i is defined as $y_i = d_i y_i^*$. The hurdle model's log-likelihood function is

$$\log(L) = \sum_{y_i=0}\left[log\left\{1 - \Phi\left(z_i\gamma, \ \frac{x_i\beta}{\sigma}\right)\right\}\right] + \sum_{y_i>0}\left[log\left\{\Phi\left(z_i\gamma + \frac{1}{\sigma}(y_i - x_i\beta)\right)\right\}\right] \ .$$

The double hurdle model has been utilized by variety topics and fields, then many researchers have attempted to modify and expand of the model's specification and the statistical properties. Recently, we can apply the Hurdle model to both count and continuous outcome, also panel data ((Dong and Kaiser (2004), Zhang, Huang, Lin and Epperson (2008)). Engel and Moffat (2014) relax the correlation assumption between two equations in both cross section and panel settings.

5. Empirical Analysis

5.1 Data Descriptions

First of all, we introduce our target variables briefly. Our dependent variables are the count process data and the monetary based quantity data. The count data is the number of revisits of each customer during the period. The monetary based outcomes are the average amount of payment and the total payment of each visit of each customer. Because we associate last visit covariate with current visit, almost all covariates are lagged variables other than "gender" and "distance." Adopting the lagged variables can avoid the endogeneity problem among covariates and error term. Table 6 and A1. show the explanations of the covariates and summary statistics, respectively. "age" is the customer's age on the day of their last visit, "$age^2/1000$" is squared term of age. "gender" is coded as 0 for male and 1 for female. "distance" indicates the distance from a customer's residence to the salon. These three are demographic variables.

As a next, we introduce the variables of customer's behavior in the last visit. "total payment" is amount of total payment and "discount rate" is receiving discount rate. When customers make an appointment with a specific hairdresser by phone call, it is "appointment" =1 and otherwise is 0. Making an appointment suggests that customers satisfied with the hairdressers' skill in the last visit. We regard this appointment as proxy variables of hairdresser's skill driven customer loyalty. "hair products" is dummy variable, if a customer purchase hair styling and/or hair care products, "hair products"=1 otherwise is 0. During the service, hairdresser recommend a hair care product to the customer, then some of customers follow the recommendation. However, the hair salon sells the exclusive professional products, the prices are higher than common toiletry in the market. The average price of hair product and hair service are 3700 yen and 4700 yen, respectively. We consider hair products as proxy variables for representing reliable or loyalty against the hair salon's product quality or their own hairdressers. Therefore, we regard "appointment" as proxy variable for representing reliable or loyalty against skills of their own hairdressers. Also, we regard "hair products" as proxy variable for representing reliable or loyalty against hair salon products' quality and/ or their own hairdressers' recommendations. We create the dummy variables concerned with the received treatments and there are four combinations that exist such as 1) “haircuts,” 2) “hair coloring with haircuts,” 3) “permanent wave with haircuts,” and 4) “haircuts, coloring and permanent wave”. These six variables represent the customer experience based on their decisions and behavior at the hair salon. Finally, we introduce four variables related supply side. “hair salon’s congestion” is presents the level of congestion in last time visits. In order to measure “hair salon’s congestion”, we divide the daily number of customers by the 90 percentile of number of customers over the year. We assume the customers may pay attention to the congestion not only of the hair salon as a whole, but also that of their own hair hairdresser. “hairdresser’s congestion” is the daily number of customers for each hairdresser. “hairdresser's fatigue” is examined by log of the receipt number of each hairdresser. We assume when the receipt number is larger, the hairdresser is probably getting tired. These three variables are unobservable and uncontrollable for the customers, and we think these experiences based on environment and atmosphere of the hair salon might affect the next revisit behaviors.

Moreover, we control the time variant skill/experiences of hairdressers by “hairdresser’s skill and capacity”. It is defined by the 90 percentile of number of customers in every year for the hairdressers. The hairdresser's capacity variable is proxy for each hairdresser's skill or productivity year by year. Using these supply side variables, we could control unobservable time-varying heterogeneity among the hairdressers. We are interested in the influences of these variables on each customer's attitude. Additionally, using the hairdresser's dummy variables as “hairdresser_#”, we can control the time invariant heterogeneity among hairdressers.

Table 6. Data Descriptions and expected sign

variable name	Description	expected sign
age	age of customers on the last visit day	+ or -
age^2/1000	(age*age/1000)	+ or -
gender	customer's gender, male=0 and female=1	+ or -
distance	distance between customer's residence and the hair salon	-
discount rate	(each service' price/total amount of payment)*100	+ or -
appointment (dummy vars)	dummy variable, No appointment=0, with appointment=0	+
hair product (dummy vars)	purchasing hair product or taken special care=1, otherwise=0	+
hair cut (dummy vars)	taken hair cut=1, otherwise=0	+ or -
hair color with hair cut (dummy vars)	taken hair color with hair cut=1, otherwise=0	+ or -
permanent wave with hair cut (dummy vars)	taken permanent wave with hair cut=1, otherwise=0	+ or -
hair cut, color and permanent wave (dummy vars)	purchase hair product=1, otherwise=0	+ or -
hair salon's congestion	# of customers (daily)/90 percentile of # of customers (yearly)	+ or -
hairdresser's congestion	count # of customers for each hairdresser, daily based data	+ or -
hairdresser's fatigue	log of (ith customer/total # of customer) for each hairdresser,	-
skill / capacity of hairdressers	estimate 90percentile of # of customers for each hairdresser, annual based data	+
hairdressers_1-11 (dummy vars)	hairdressers in charge for the last visit	+ or -
spring_dummy	March, April and May=1, otherwise=0	+ or -
summer_dummy	June, July and August=1, otherwise=0	+ or -
fall_dummy	September, October and November=1, otherwise=0	+ or -
winter_dummy	December, January and February=1, otherwise=0	+ or -

Notes: We specially focus on the shaded areas' results.

5.2 Estimation Models and Results for Count Data

Table 7 provides the results of Poisson, ZIP and ZINB models for number of revisit data. Each regression equation is shown as below ((i), (ii) and (iii)). For Poisson estimation, we utilize the number of revisit data during the period for each customer i as dependent variable. We associate the first visit covariate (e.g. age_{i0}) with dependent variable. Both ZIP and ZINB model have two separate equations, one is selection model which expresses by logit equation if a customer revisits or not. Another models are Poisson model or negative binomial model for count data, when number of revisits are larger than 0. We use same dependent variable and covariates among different three estimation methods. According to the results of vuong test in ZIP estimation and ZINB estimation, ZIP model is preferred to Poisson model and ZINB is preferred to NB model, respectively. Moreover, in ZINB model, we can observe $ln(\alpha)$ is a significantly positive, it suggests there exists overdispersion in outcome, thus we adopt the ZINB's results here. In the selection logit model, we have observed the effect of the information at the time of the initial visit on the probability of revisit. Comparing the estimation results with the expected signs in Table 6, having different signs than expected signs among the significant coefficients are distance (+), purchase hair products (-), skill of hairdresser (-). For revisited customer, we observe the impact of initial visit on the total number of revisits in count model estimation. Comparing the estimation results with the expected signs, only "purchase hair products" and "skill of hairdresser" is significant in focusing variables (see Table 6). We observed a few significant coefficients, but the significant coefficients have same signs of expected signs. As a results, the Effect of initial visit gave only small impact on total number of revisits. It is considered that the initial impact has been attenuated with time and customer has changed the preference and decision after purchasing the services (Zeithaml (1981)). In this framework, we made a cross section data to compress the panel data while discarding information of time direction data. To the next section, we estimate the model by using panel data.

(i) Poisson regression:

$$\begin{aligned}\log\mu_i = c + \beta_1 age_{i0} + \beta_2 \frac{age_{i0}}{1000} + \beta_3 gender_{i0} + \beta_4 distance_{i0} + \beta_5 appoint_{i0} + \beta_6 product_{i0} \\ + \beta_7 cut_{i0} + \beta_8 corol\_cut_{i0} + \beta_9 perm\_cut_{i0} + \beta_{10} cutcolorperm_{i0} \\ + \beta_{11} hairsalon\_busy_{i0} + \beta_{12} hairdresser\_busy_{i0} + \beta_{13} hairdresser\_tired_{i0} \\ + \beta_{14} hairdresser\_skill_{i0} + \beta_{15} season_{i0}, \qquad i = 1, \cdots, \mathrm{N}\end{aligned}$$

,where subscript 0 denotes the first visit and μ_i denotes the number of revisit for each customer .

(ii) Zero-inflated Poisson:
-selection model (logit model)

$$\begin{aligned}\ln\left(\frac{1-p}{p}\right) = b + \gamma_1 age_{i0} + \gamma_2 \frac{age_i 0}{1000} + \gamma_3 gender_{i0} + \gamma_4 distance_{i0} + \gamma_5 appoint_{i0} \\ + \gamma_6 product_{i0} + \gamma_7 cut_{i0} + \gamma_8 corol\_cut_{i0} + \gamma_9 perm\_cut_{i0} \\ + \gamma_{10} cutcolorperm_{i0} + \gamma_{11} hairsalon\_busy_{i0} + \gamma_{12} hairdresser\_busy_{i0} \\ + \gamma_{13} hairdresser\_tired_{i0} + \gamma_{14} hairdresser\_skill_{i0} + \gamma_{15} season_{i0}\end{aligned}$$

-Poisson model, $y > 0$

$$\log\mu_i = c + \beta_1 age_{i0} + \beta_2 \frac{age_{i0}}{1000} + \beta_3 gender_{i0} + \beta_4 distance_{i0} + \beta_5 appoint_{i0} + \beta_6 product_{i0} + \beta_7 cut_{i0} + \beta_8 corol\_cut_{i0} + \beta_9 perm\_cut_{i0} + \beta_{10} cutcolorperm_{i0} + \beta_{11} hairsalon\_busy_{i0} + \beta_{12} hairdresser\_busy_{i0} + \beta_{13} hairdresser\_tired_{i0} + \beta_{14} hairdresser\_skill_{i0} + \beta_{15} season_{i0}$$

(iii) Zero-inflated binomial:

-selection model (logit model)

$$\ln\left(\frac{1-p}{p}\right) = ln(\alpha) + b + \gamma_1 age_{i0} + \gamma_2 \frac{age_{i0}}{1000} + \gamma_3 gender_{i0} + \gamma_4 distance_{i0} + \gamma_5 appoint_{i0} + \gamma_6 product_{i0} + \gamma_7 cut_{i0} + \gamma_8 corol\_cut_{i0} + \gamma_9 perm\_cut_{i0} + \gamma_{10} cutcolorperm_{i0} + \gamma_{11} hairsalon\_busy_{i0} + \gamma_{12} hairdresser\_busy_{i0} + \gamma_{13} hairdresser\_tired_{i0} + \gamma_{14} hairdresser\_skill_{i0} + \gamma_{15} season_{i0}$$

, $ln(\alpha)$ represents an exposure term.

-negative binomial model, $y > 0$

$$\log\mu_i = ln(\alpha) + c + \beta_1 age_{i0} + \beta_2 \frac{age_{i0}}{1000} + \beta_3 gender_{i0} + \beta_4 distance_{i0} + \beta_5 appoint_{i0} + \beta_6 product_{i0} + \beta_7 cut_{i0} + \beta_8 corol\_cut_{i0} + \beta_9 perm\_cut_{i0} + \beta_{10} cutcolorperm_{i0} + \beta_{11} hairsalon\_busy_{i0} + \beta_{12} hairdresser\_busy_{i0} + \beta_{13} hairdresser\_tired_{i0} + \beta_{14} hairdresser\_skill_{i0} + \beta_{15} season_{i0}$$

5.3 Estimation Models and Results for Quantity Data

We adopt two outcomes, first one is total amount of purchase of each customer by each day (hereafter, total payment), another is cumulative sum of total payment divide by cumulative sum of number of visit for each customer by each day (hereafter, average payment). The total payment model tells us which factor has impact on the demand of hair services. The average payment implies that the growth of loyalty or the degree of patronage. To estimate these two models, we utilize a double hurdle model which allows for different specification of the participation (revisit or non-revisit) and consumption processes and the possible correlation between these two processes. The hurdle model concerns that the unbounded outcomes are over jump a hurdle. The double hurdle can be expressed by combination of probit for participation model and tobit estimator for purchasing model. We adopt covariates x and z as same as the count model in 5.2 other than "total payment" and "discount rate". The covariates are observed on the last visit.

Double hurdle model

-Probit model-

$$d_i^* = z_i' \gamma + \varepsilon_{1,i}$$
$$d_i = 1 \ \ if \ \ d_i^* > 0$$
$$d_i = 0 \ \ if \ \ d_i^* \leq 0.$$

-Tobit model

$$y_{it}^* = x_{it-1}' \beta + u_{it}$$

$$y_i = \begin{cases} y_{it}^* & if \quad y_{it}^* \geq 0 \\ 0 & if \quad y_{it}^* < 0 \end{cases}$$

The estimated coefficients indicate the marginal effect of x on expectation of latent variable y_{it}^*. In original Cragg (1971) model, it assumes that the error terms of two equations are uncorrelated in eq. (5). We also implement estimation proposed by Engel and Moffat (2014) which relaxed the correlation assumption between two equations. Table 8 and 9 show the estimation results for total payment model and average payment model, respectively. In both Tables, we list the estimation results of "Independent" and "Correlation" which are assumed whether error terms are independent or not. When the "mills ratio" is significant, $H_0\colon Corr(\varepsilon_1 \varepsilon_2) = 0$ is rejected. In our results, both mills ratios are insignificant, it suggests that we don't need care about the correlation between probit and tobit models. However, as mentioned in 4.1, in hair salon or restaurant case, once customers enter the shop, they will buy the services or product certainly at the same time. Therefore, we have adopted as the

estimation results that considering the correlation between two models.
Among two models in Table 8 and 9, we observed age coefficients are significantly positive and squared of age are significantly negative both of consumption and participation models. If age has a positive coefficient, it may imply that older customers have enough money to care for and maintain their hairstyle as compared to the younger customers. In a certain case, age might be regarded as a proxy variable for income in one possibility. The coefficients of gender are significant at 5% level, but the signs are different between participation and consumption model. Since gender is coded as 0 for male and 1 for female, negative coefficients in participation models imply that male customers have a higher probability of revisit compared to female customers. On the other hand, the positive coefficients of consumption models indicate that female customers purchase more than male customers. The coefficients of distance show always strongly negative. It implies that the customers who live further away will visit the salon later (or not revisit) and their payments are less. As a next step, we will explain about the coefficients of appointment. Appointment=1 is when a customer makes an appointment with a specific hairdresser, and 0 denotes other choices. The coefficients of appointment have a strong positive impact on the both models in Table 8 and 9. We may say that the customer who makes an appointment with their favorite hairdresser will come back to the salon and purchase more than the customer who does not. We regard this appointment as proxy variables of customer loyalty which is originated by hairdresser's skill. We also consider hair products as proxy variables for representing reliable or loyalty against the quality of hair products and/or their own hairdressers' recommendations. "hair products" is dummy variable, if a customer purchase hair styling and/or hair care products, hair products=1 otherwise is 0. The coefficients of hair products are significantly positive. The customer who has higher loyalty will revisit with higher probability and they will spend more money in the next visit. We create the dummy variables concerned with the received treatments and there are four combinations that exist such as 1) "haircuts," 2) "hair coloring with haircuts," 3) "permanent wave with haircuts," and 4) "haircuts, coloring and permanent wave". These four variables are distributed as customers' choice variables. The base combination that is customer who taken hair color, permanent wave, shampoo and other hair care services without haircut. In 1) haircuts case, the coefficients are significantly positive in participation regressions, it presents that customers who only had a haircut on their first has higher

probability of revisits than any other combinations. Also, the negative coefficients in purchase regression mean that the customer pay less than customers who had other services. This estimation results for hair service dummy variables are consistent with Table 5. That's are behavioral variables.

From now, we show the results of supply side and environment variables. "hair salon's congestion" is presents the level of congestion in last time visits. The coefficients are significantly negative in the participation regressions, it suggests that the salon's impression is impulsive, then it discourage willingness of customer's revisit. On the contrary, the purchase models have positive coefficients of hair salon's congestion. It indicates that the customers who revisit multiple time prefer the active salon. We assume the customers may pay attention to the congestion not only of the hair salon as a whole, but also that of their own hair hairdresser. hairdresser's congestion is the daily number of customers for each hairdresser. The coefficients are insignificant in the purchase regressions and significantly positive, but the magnitudes are small, in the revisit probability regressions. "hairdresser's fatigue" is examined by log of the receipt number of each hairdresser. We assume when the receipt number is larger, the hairdresser is probably getting tired. We found the coefficients are strongly significant and the signs are negative in all models. The variable may capture of hairdressers' conditions, feeling and some atmosphere of the memorable for customers. These three variables are unobservable and uncontrollable for the customers, and we think these experiences based on environment and atmosphere of the hair salon might affect the next revisit and purchase behaviors. Finally, we control the time variant skill/experiences of hairdressers by "hairdresser's skill and capacity". The hairdresser's capacity variable is proxy for each hairdresser's skill or productivity year by year. Using these variables, we could control unobservable time-varying heterogeneity among the hairdressers. We observe the positive coefficients of hairdressers' skill in all models. We can say the skill or productivity of hairdressers contribute to raise the probability of revisit, total amount of sales and the loyalty. We also adopt hairdressers dummy variables to control time-invariant heterogeneity effects among the eleven hairdressers.

Table 10 is reused of Table 8 and 9's results focusing on important six variables. In this study, we challenged to reproduce the experience space of customers at hair salon by using the customer behavioral variables, loyalty data, the hairdresser's skill and hair salon's atmosphere or environment in addition to demographic variables. These new adopted variables influence the customer's attitude toward the revisiting behavior and purchase decisions at the salon. Comparing the estimation results of revisit probability, number of revisits and purchase regressions, we found that customer’s purchase decisions or behaviors are more sensitive about the salon's atmosphere, congestion, hairdressers' skills and their own experiences at the salon. We can also indicate the important role of functional quality as well as technical quality.

Table 7. Estimation Results for Count Data, Dependent variable: number of revisits

	normal poisson	Zero-Inflated Poisson		Zero-Inflated Negative Binomial	
covariates	count Poisson model	count Poisson model	selection logit model	count Negative Binomial	selection logit model
age	-0.005	-0.015	-0.031***	-0.015	-0.097***
	(0.015)	(0.012)	(0.011)	(0.011)	(0.034)
age^2/1000	0.176	0.304	0.398**	0.285*	1.281***
	(0.223)	(0.191)	(0.167)	(0.167)	(0.467)
gender	-0.302***	-0.275***	0.070	-0.344***	-0.360
	(0.052)	(0.047)	(0.064)	(0.057)	(0.228)
distance	-0.006***	-0.003*	0.008***	-0.004***	0.015***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.004)
total payment	0.033***	0.007	-0.063***	0.011	-0.236***
	(0.007)	(0.006)	(0.009)	(0.008)	(0.037)
discount rate	0.017***	0.009***	-0.018***	0.012***	-0.038***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.006)
appointment	0.098***	-0.030	-0.313***	-0.010	-1.008***
	(0.036)	(0.031)	(0.041)	(0.037)	(0.205)
hair product	0.199***	0.104***	-0.241***	0.159***	-0.420**
	(0.043)	(0.039)	(0.055)	(0.046)	(0.212)
hair cut	0.382***	-0.002	-0.838***	0.056	-2.219***
	(0.079)	(0.074)	(0.069)	(0.093)	(0.275)
hair color with hair cut	0.170**	-0.024	-0.362***	0.018	-0.309
	(0.071)	(0.064)	(0.066)	(0.081)	(0.200)
parmanent wave with	0.070	-0.070	-0.233***	-0.058	-0.078
hair cut	(0.076)	(0.068)	(0.074)	(0.086)	(0.221)
haircut, hair color and	0.226**	0.081	-0.253**	0.106	0.347
parmanent wave	(0.103)	(0.092)	(0.124)	(0.110)	(0.455)
hair salon's congestion	-0.138*	-0.115*	0.014	-0.114	-0.320
	(0.078)	(0.069)	(0.092)	(0.089)	(0.319)
hair dresser's	-0.001	0.001	0.006*	0.001	0.026**
congestion	(0.003)	(0.002)	(0.003)	(0.003)	(0.011)
hair dresser's	-0.041*	-0.028	0.028	-0.040	-0.051
fatigue	(0.024)	(0.021)	(0.027)	(0.025)	(0.082)
skill/capacity of	0.010***	0.004	-0.014***	0.006*	-0.032***
hairdressers	(0.003)	(0.002)	(0.003)	(0.003)	(0.011)
spring_dummy	-0.047	-0.008	0.084	-0.041	0.133
	(0.050)	(0.045)	(0.055)	(0.054)	(0.178)
summer_dummy	0.164***	0.120***	-0.083	0.142***	0.078
	(0.046)	(0.041)	(0.052)	(0.050)	(0.166)
fall_dummy	0.013	-0.001	-0.046	-0.018	-0.165
	(0.046)	(0.041)	(0.052)	(0.049)	(0.173)
Constant	0.656***	2.067***	1.727***	1.567***	3.911***
	(0.253)	(0.222)	(0.227)	(0.221)	(0.748)
ln(alpha)				**0.810*** (c)**	
				(0.030)	
AIC	9.62	6.87		4.41	
vuong test stat.		42.62*** (a)		8.53*** (b)	
Observations	13,315	13,315		13,315	

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
(a) normal poisson versus zero-inflated poisson, vuong statistics of 42.62 indicates that zero-iflated poisson is favored. (b) negative binomial versus zero-inflated n.b., vuong statistics of 8.53 indicates that zero-iflated n.b. is favored. (c) ln(alpha) is significant indicates that zero-inflated n.b. is prefered to zero-inflated poisson model.

Table 8. Estimation Results for Quantity Data: Total Payment by day (Yen)

	Independent		Correlation	
covariates	Consumption	Participation	Consumption	Participation
age	187.451***	-0.001	193.507***	0.016**
	(11.868)	(0.007)	(11.651)	(0.007)
age^2/1000	-1,966.825***	0.023	-2,047.755***	-0.117
	(156.837)	(0.092)	(152.155)	(0.092)
gender	2,923.316***	-0.695***	2,800.957***	-0.647***
	(69.403)	(0.085)	(71.437)	(0.084)
distance	-0.547	-0.010***	-3.126	-0.008***
	(2.106)	(0.001)	(2.175)	(0.001)
appointment	1,067.929***	1.744***	1,190.713***	1.647***
	(65.630)	(0.030)	(205.906)	(0.029)
hair product	1,509.357***	0.061**	1,556.871***	0.068**
	(59.092)	(0.031)	(59.139)	(0.031)
hair cut	-1,378.620***	0.491***	-1,305.575***	0.511***
	(77.607)	(0.046)	(80.867)	(0.048)
hair color with hair cut	1,242.050***	0.003	1,370.103***	-0.032
	(74.140)	(0.036)	(73.826)	(0.036)
permanent wave with	309.449***	-0.115***	394.844***	-0.087**
hair cut	(85.600)	(0.040)	(85.969)	(0.041)
haircut, hair color and	2,229.078***	0.018	2,186.491***	0.022
permanent wave	(115.885)	(0.058)	(115.359)	(0.059)
hair salon's congestion	367.561***	-0.267***	434.795***	-0.251***
	(114.743)	(0.059)	(114.539)	(0.058)
hair dresser's	3.808	0.002	0.044	0.006***
congestion	(3.701)	(0.002)	(3.672)	(0.002)
hair dresser's	-134.892***	-0.051***	-158.626***	-0.075***
fatigue	(32.261)	(0.018)	(32.129)	(0.018)
skill/capacity of	65.304***	0.003	67.012***	0.003*
hairdressers	(5.990)	(0.002)	(5.950)	(0.002)
spring_dummy	102.958	-0.104***	193.678***	-0.101***
	(63.061)	(0.036)	(63.063)	(0.036)
summer_dummy	308.599***	-0.114***	367.922***	-0.075**
	(62.443)	(0.034)	(62.293)	(0.034)
fall_dummy	151.555**	-0.081**	182.280***	-0.102***
	(62.028)	(0.034)	(62.230)	(0.034)
mills ratio			318.627	
			(432.713)	
σ	4,769.741***		4,751.879***	
	(17.213)		(17.128)	
Constant	-700.839**	1.382***	-342.058	0.978***
	(357.470)	(0.158)	(419.176)	(0.155)
Observations	55,821	55,821	55,821	55,821

Notes: Bootstrapped Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
hairdresser dummy variables are included.

Table 9. Estimation Results of Quantity Data: Average Payment (Yen)

	Independent		Correlation	
covariates	Consumption	Participation	Consumption	Participation
age	135.453***	0.031***	138.076***	0.041***
	(5.813)	(0.004)	(6.821)	(0.004)
age^2/1000	-1,405.373***	-0.354***	-1,437.175***	-0.433***
	(77.142)	(0.060)	(86.747)	(0.061)
gender	2,046.947***	-0.200***	2,053.757***	-0.204***
	(34.771)	(0.034)	(36.974)	(0.034)
distance	-1.032	-0.009***	-3.667***	-0.008***
	(1.025)	(0.001)	(1.234)	(0.001)
appointment	537.713***	1.391***	490.900***	1.325***
	(30.093)	(0.017)	(182.247)	(0.017)
hair product	1,539.851***	0.124***	1,503.096***	0.146***
	(28.727)	(0.024)	(30.915)	(0.024)
hair cut	-1,807.931***	0.207***	-1,795.306***	0.204***
	(37.305)	(0.029)	(41.258)	(0.029)
hair color with hair cut	1,283.340***	0.077***	1,292.744***	0.046*
	(35.337)	(0.026)	(36.027)	(0.026)
permanent wave with	1,030.774***	-0.071**	1,079.124***	-0.065**
hair cut	(40.726)	(0.029)	(41.709)	(0.029)
haircut, hair color and	3,188.948***	0.138***	3,197.359***	0.135***
permanent wave	(55.877)	(0.045)	(57.942)	(0.046)
hair salon's congestion	152.117***	-0.139***	176.284***	-0.126***
	(55.775)	(0.042)	(56.523)	(0.042)
hair dresser's	2.701	0.001	0.779	0.003**
congestion	(1.798)	(0.001)	(1.804)	(0.001)
hair dresser's	-66.455***	-0.046***	-91.484***	-0.067***
fatigue	(15.626)	(0.013)	(16.172)	(0.013)
skill/capacity of	36.278***	0.006***	35.499***	0.007***
hairdressers	(2.926)	(0.001)	(2.969)	(0.001)
spring_dummy	41.728	-0.072***	131.496***	-0.065***
	(30.690)	(0.025)	(31.300)	(0.025)
summer_dummy	145.533***	-0.048**	226.723***	-0.018
	(30.395)	(0.024)	(30.714)	(0.024)
fall_dummy	49.337	-0.034	109.736***	-0.046*
	(30.213)	(0.024)	(30.689)	(0.024)
mills ratio			-72.673	
			(385.985)	
σ	2,370.622***		2,383.109***	
	(7.706)		(7.732)	
Constant	2,452.036***	0.037	2,476.698***	-0.172*
	(174.423)	(0.094)	(327.163)	(0.093)
Observations	55,821	55,821	55,821	55,821

Notes: Bootstrapped Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
hairdresser dummy variables are included.

Table 10. Estimation Results (Reused, Table 8 and 9)

variables	total payment by day		average payment		expected sign	Label
	consumption	participation	consumption	participation		
appointment (*1)	1,190.713***	1.647***	490.900***	1.325***	+	Loyalty
	(205.906)	(0.029)	(182.247)	(0.017)		
hair products (*2)	1,556.871***	0.068**	1,503.096***	0.146***	+	
	(59.139)	(0.031)	(30.915)	(0.024)		
hair salon's congestion	434.795***	-0.251***	176.284***	-0.126***	+ or -	Functional Quality
	(114.539)	(0.058)	(56.523)	(0.042)		
hair dresser's congestion	0.044	0.006***	0.779	0.003**	-	
	(3.672)	(0.002)	(1.804)	(0.001)		
hair dresser's fatigue	-158.626***	-0.075***	-91.484***	-0.067***	-	
	(32.129)	(0.003)	(16.172)	(0.013)		
skill/capacity of hairdressers	67.012***	0.003*	35.499***	0.007***	+	Technical Quality
	(5.950)	(0.002)	(2.969)	(0.001)		

(*1) We regard "appointment" as proxy variables of customer loyalty which is originated by hairdresser's skill.
(*2) We regard "hair products" as proxy variables of customer loyalty based on hair products' quality of hair salon.

6. Conclusion

What do hair salon and hairdressers provide customers? One answer is that they provide services and create customer satisfaction. The customer satisfaction could depend on factors such as location, price, the skill of the hairdresser, and the overall experience at the hair salon. It is difficult to observe customer satisfaction directly for hair salon owners and researchers, because satisfaction is subjective. However, if the customers like the hair salon or hairdressers and are satisfied with their services, customers would come back to the hair salon, come back more often and become high-spender. This suggests that we can measure customer satisfaction by investigating what causes them to come back. In this study, we apply count process and double hurdle analysis to specify a model of customers' return and purchase behavior. We collected a daily record from a hair salon in Japan between 2003 to 2010 (2,046 working days). The hair salon has about 15,000 customers and the daily records hold payments and hairdresser's names for every customer's treatment.

In this study, we challenged to reproduce the experience space of customers at hair salon by using the customer behavioral variables, loyalty data, the hairdresser's skill and hair salon's atmosphere or environment in addition to demographic variables. These new adopted variables influence the customer's attitude toward the revisiting behavior and purchase decisions at the salon. Comparing the estimation results of revisit probability, number of revisits and purchase regressions, we found that customer's purchase decisions or behaviors are more sensitive about the salon's atmosphere, congestion, hairdressers' skills and their own experiences at the salon. We can also indicate the important role of functional quality as well as technical quality.

Finally, we show the limitation of this research was conducted in one hair salon in Japan, thus may not be generalizable to other hair salons. Therefore, more salons need to be investigated. It is also suggested that other related factors of service quality such as internal marketing, detailed hairdresser's quality and skill, peer review among workers must be helpful to measure their productivity and quality. These will be included in the future research.

7. References

1. Cragg, J. G. (1971) "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica* 39, pp.829-844.

2. Cronin J. J., and Taylor S. A. (1992) "Measuring Service Quality: A Reexamination and Extension," *Journal of Marketing* 56, pp. 55-68.

3. Darby, M.R. and E. Karni (1973) "Free Competition and the Optimal Amount of Fraud," *Journal of Law and Economics* 16, pp. 67-86.

4. Dong, D., and H. M. Kaiser (2004) "Modelling Milk Purchasing Behaviour with a Panel Data Double Hurdle Model," *Applied Economics* 36, pp. 769-779.

5. Elickson, P. B. (2007) "Does Sutton Apply to Supermarkets?" *The RAND Jounral of Economics* 38, pp. 43-59.

6. Engel, C. and Moffat, P. G. (2014) "dhreg, xtdhreg, and bootdhreg: Commands to Implement Double-Hurdle Regression," The STATA Journal 14, pp. 778-797.

7. Grönroos, C. (1988) "Service Quality: The Six Criteria of Good Perceived Service Quality," *Review of Business* 9, pp. 10-13.

8. Gross, J. J. (2015) “Emotion Regulation: Current Status and Future Prospects,” *Psychological Inquiry* 26, PP. 1-26.

9. Hussian, R. and Ali, M. (2015) “Effect of Store Atmosphere on Consumer Purchase Intention,” *International Journal of Marketing Studies* 7, pp. 35-43.

10. Jones, M. A., Mothersbaugh, D. L., and Beatty, S.E. (2000) “Switching Barriers and Repurchase Intention in Services,” *Journal of Retailing* 76, pp259-274.

11. Kolko, J. (2007) “Urbanization, Agglomeration, and Coagglomeration of Service,” in: Glaeser, E. L. (Eds.): Agglomeration Economics, The University of Chicago Press, pp. 151-180.

12. Konishi, Y. (2010) “The Effects of Congestion and Skills at a Hair Salon on the Consumer's Revisiting Behavior,” *RIETI Discussion Paper*, 110-E-46.

13. Konishi, Y. and Y. Nishiyama (2010) “Productivity of Service Providers: Microeconometric measurement in the case of hair salons,” *RIETI Discussion Paper*, 10-E-51.

14. Morikawa, M. “Economies of Densities and Productivity in Service Industries: An Analysis of Personal-Service Industries Based on Establishment Level Data,” Review of Economics and Statistics 93, pp. 179-192.

15. Nelson, P. (1970) “Information and Consumer Behavior,” *Journal of Political Economy* 78, pp. 311-329.

16. Oi, W. (1992). “Productivity in the Distributive Trades: The Shopper and the Economics of Massed Reserves”, in Z. Griliches (Eds.) Output Measurement in the Service Sector, Chicago: University of Chicago Press.

17. Oliver, R. (1999). “Whence Customer Loyalty?”, *Journal of Marketing* 63 (Special Issue), 33-44.

18. Oliver, R. (1993) “Cognitive, Affective and Attribute Bases of Satisfaction Response,” *Journal of Consumer Research* 20, pp. 178-186.

19. Parasuraman, A., Zeithaml, V., & Berry, L. (1988), "SEVQUAL: A multiple-item scale for measuring consumer perceptions of service quality", *Journal of Retailing*, Vol. 64, pp.12-37.

20. Schmenner, R.W. (1986), "How can Service Businesses Survive and Prosper?", *Sloan Management Review*, Vol. 27 No. 3, pp. 21-32

21. Zeithaml, V. A. (1981) "How Consumer Evaluation Process Differ Between Goods and Services," in: Donnelly, J. and George, W. (Eds.): Marketing of Services: Proceedings of the American Marketing Association Conference, Chicago, pp. 186-190.

22. Zeithaml, V. A., Berry, L. L. and Parasuraman, A. (1996) "The Behavioral Consequences of Service Quality," *Journal of Marketing* 60, pp. 31-46.

23. Zhang, F., Huang, C. L., Lin, B. H. and Epperson, J. E. (2008) "Modeling Fresh Organic Produce Consumption with Scanner Data: A Generalized Double Hurdle Model Approach," *Agribusiness* 24, PP. 510-522.

Appendix

Table A1. Summary Statistics

variables	mean	median	std.dev.	min	max	Obs
num of visit times (*)	8.015	5	10.12	1	147	63,226
total paymen by day (*)	8,245	8,400	5,287	0	42,320	63,226
average payment (*)	7,820	8,190	3,996	0	29,820	63,226
age	28.21	26	8.198	10	84	63,226
age^2/1000	0.863	0.676	0.608	0.100	7.056	63,226
gender	0.875	1	0.330	0	1	63,226
distance	8.680	5.200	10.56	0.0600	97.89	63,226
discount_rate	13.37	16.07	11.27	0	100	63,193
nob of purchased services	2.312	2	1.152	1	13	63,226
appointment (dummy vars)	0.790	1	0.407	0	1	63,226
hair product (dummy vars)	0.164	0	0.370	0	1	63,226
hair cut (dummy vars)	0.302	0	0.459	0	1	63,226
hair color with hair cut (dummy vars)	0.356	0	0.479	0	1	63,226
permanet wave with hair cut (dummy vars)	0.160	0	0.366	0	1	63,226
hair cut, color and permanent wave (dummy vars)	0.0513	0	0.221	0	1	63,226
hair salon's congestion	0.746	0.724	0.247	0.00855	1.538	63,226
hairdresser's congestion	20.41	19	10.62	1	57	63,226
hairdresser's fatigue	1.398	1.386	0.821	0	3.296	63,226
skill / capacity of hairdressers	27.54	28	9.093	1	39	63,226
hairdressers_1 (dummy vars)	0.286	0	0.452	0	1	63,226
hairdressers_2 (dummy vars)	0.0997	0	0.300	0	1	63,226
hairdressers_3 (dummy vars)	0.0939	0	0.292	0	1	63,226
hairdressers_4 (dummy vars)	0.0384	0	0.192	0	1	63,226
hairdressers_5 (dummy vars)	0.0802	0	0.272	0	1	63,226
hairdressers_6 (dummy vars)	0.176	0	0.381	0	1	63,226
hairdressers_7 (dummy vars)	0.0967	0	0.296	0	1	63,226
hairdressers_8 (dummy vars)	0.0233	0	0.151	0	1	63,226
hairdressers_9 (dummy vars)	0.0582	0	0.234	0	1	63,226
hairdressers_10 (dummy vars)	0.0400	0	0.196	0	1	63,226
spring_dummy	0.243	0	0.429	0	1	63,226
summer_dummy	0.249	0	0.432	0	1	63,226
fall_dummy	0.258	0	0.437	0	1	63,226
winter_dummy	0.250	0	0.433	0	1	63,226

(*) indicates that the variables are dependent variable in Section 5.