RIETI Discussion Paper Series 10-E-046

The Effects of Congestion and Skills at a Hair Salon on the Consumer's Revisiting Behavior

KONISHI Yoko
RIETI

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

RIETI Discussion Paper Series 10-E-046
September 2010

The Effects of Congestion and Skills at a Hair Salon on the Consumer's Revisiting Behavior

Y. Konishi
(Fellow, RIETI; Japan
JSPS Postdoctoral Fellowships for Research Abroad; Yale University)

Abstract

In this study, we apply duration analysis to specify a model of customers' behavior regarding their visits to a hair salon and we then estimate their aggregated revisit rates / dates. We adopt the following approach: [1] assuming that there are differences among first time customers, regular and loyal customers with respect to the intensity function; [2] introducing customers' behavioral variables, hair salon congestion and hairdresser skill variables, in addition to demographic variables; [3] by applying the estimation results of Cox regression, we examine the aggregated revisit rate and show how we measure the individual next revisit date. As a result, we found that the intensity functions of non-loyal and loyal customers have been specified by different models. We could observe differences between first time customers and loyal customers in terms of the response toward hair salon and hairdresser congestion. We could also find that loyal customers tend to prefer higher skilled hairdressers. It suggests that customer heterogeneity should be included in the intensity model and that we also need the hair salon's information (supply side) to specify customers' revisit model.

Key words: Demand Structure; Consumer Satisfaction; Individual Service Industry; Hair Salon
JEL classification: D12, D22, L84, M31

The author would like to thank RIETI and JSPS for their support. I am grateful to Kazuo Yamaguchi, Kotaro Tsuru and seminar participants at RIETI for their helpful comments and suggestions.

1. Introduction

Recently, with the increasing importance and need of one-to-one marketing in some industries, e.g., retail, restaurant, hotel industries and so on, direct marketing has modified its contents and concept. It might imply that we should consider heterogeneity for each individual and each industry. In the beauty industry, direct mail has apparently been used as one of the marketing strategies to contact customers and provide them with incentives to revisit their salon. However, in terms of sending direct mail, salons might have dealt with the customers in a similar manner. Direct mailing could probably serve as a marketing strategy for the salon to obtain more profits. In order to realize the optimal direct mailing strategy, it would be necessary to focus on an adequate target base, that is, we should distinguish the customers who intend to patronize the salon from other customers. The segmentation problem is one of the primary issues in marketing and it has been dealt with in earlier studies. In early works, segmentation might often be recognized as a purpose in itself, and the segmentation method denotes dividing customers into homogeneous groups without use of objective variables. Recently, Jonker et al. (2004) have highlighted that optimal segmentation leads to the realization of mailing strategy optimization. Moreover they suggested the necessity of exploring systematic analytic procedures to optimize both these steps. In their study, they applied one of the segmentation methods that was introduced as a standard and useful method in several prior studies in this area - the RFM methods (Hughes, 1994). The RFM method measures the customer value based on R (Recency), F (Frequency), and M (Monetary value). It is recognized as one of the helpful methods to separate the good (more profitable) customers from the other customers in order to achieve better segmentation.

Fortunately, nearly all Japanese hair salons require a customer to fill in a questionnaire on their first visit. Salon owners can obtain data on a customer's sex, age, occupation, and address and moreover they can use the monetary data based on the customer's behavior by recording it; these data should be helpful in the identification and segmentation of each customer. In this study, we apply duration analysis to specify a model of hair salon customer behavior regarding their visits and then estimate their aggregated revisit rates/dates. The dataset of this study was obtained from a hair salon located in Japan's third largest city, Osaka, and our target customers primarily use the haircut service. We adopt the following approach: [1] assuming that there are differences among first visit customers, regular and loyal customers with respect to the intensity function; [2] introducing customers' behavioral variables, hair salon congestion and hairdresser's skill variables in addition to demographic variables; [3] observing the degree of sensitivity to the congestion and skill to intensity function by age cohorts; [4] by applying estimation results of Cox regression, we examine the aggregated revisit rate and show the procedure of measuring the individual next revisit date.

As a result, we found that the intensity functions of non-loyal and loyal customers have been specified by different models. We could observe differences between the first visit customers and loyal customers in terms of the responses against hair salon and hairdresser congestion. We could also find that loyal customers tend to prefer higher skilled hairdressers. It suggests that customer heterogeneity should be included in the intensity model and we also need the information of the hair salon (supply side) to specify customers' revisit model. Using our estimation results, we could predict revisit rates for each customer. It will tell us when and how many customers will visit a salon and which services they will receive. It should be helpful to predict daily sales, the timing of sending direct mail and optimal allocation of labor and capital.

The following section presents preliminary analysis regarding the summary of the variables through an empirical study and provides some suggestions regarding the segmentation of customers. Section 3 reviews the Cox model and the methodology in detail and provides the estimation results and also shows the procedure of prediction of individual revisit rate. The concluding remarks are given in Section 4.

2. Preliminary Data Analysis

2.1. **Description of a Hair Salon.** We obtained data regarding customer attributes and customer records from a hair salon. The hair salon is located in Japan's third largest city, Osaka. The salon is easily accessible from certain 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 sex, age, occupation, address, hobby and some other preferences. The hair salon opened in July 2003, and we observed them from July 2003 to March 2010 (2,418days). The hair salon provides the customers with various hair products and services including not only haircuts, color, and permanent waves, but also nail and facial care, makeup, and so on. As you know, their main services are haircuts, color, and permanent waves and their share of total number of treatments are 36%, 21% and 10% respectively. The three services yield more than 80% of the total sales.

In this paper, we use the customers' sex, age and the distance between the salon and their home addresses as demographic variables and also behavioral variables in our empirical study. They have 16,000 customers comprising approximately more than 90% females. The average age of customers is 29 years old (youngest is 12 and oldest 84). Thus, the majority of customers lie between 20 and 30 years of age. The distance between the residences of the customers and the salon vary greatly. We observed forty prefectures in their addresses. The mean of distance is around 10 km. Surprisingly, the maximum distance is more than 1,000 km. However, only 2% of the customers live this far from the hair salon. 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 they are treated as outliers. The remaining 98% customers are from four prefectures (Osaka, Kyoto, Hyougo and Nara) and 88% of the customers are from Osaka prefecture. After we abstracted the customers who live within 50km from the hair salon, about 3% of the customers are removed. We can use this demographic information from the initial questionnaire and also 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 their productivity. In order to specify the model of customers' behavior, we use the customers' demographic and behavioral variables, skill or experience of hairdressers and level of the hair salon's congestion variable.

2.2. **Data Construction for Our Empirical Study: making Segmentaions.** We specify the consumers' revisit behavior using the survival analysis framework. To implement the estimation, we need to obtain duration data which are the interval between the last visit and the visit prior to it. In our dataset, the end point is the 2,418th day (type 1 Censor). We treat the interval from the end point to the latest visit as the censoring data. In order to construct a target dataset for our analysis, first, we abstract from the full dataset, the data on the customers who used the haircut service and then remove the customers with the following conditions: [1] customers who live 50 km away from the hair salon; [2] customers who don't tell their birthday to the salon; [3] customers who cannot be identified as individuals, e.g. name, id number and address are missing. After abstracting our target dataset, 11,334 customers remain (full data set has 16,000 customers).

As a next step, we show how to make segmentations in our target dataset. According to previous studies, Korgaonkar et al. (1985) and Akaah et al. (1995) describe that a customer's intention (low or high) might have significantly influenced the customer's attitudes toward direct marketing. This is based only on their conjectures even though it does not appear in their empirical results. In this study, we think there are different revisit behaviors among the first visitor and regular customers. Table 1 shows the descriptive statistics of number of visits of each customer for seven years.

In order to divide customers into segmentations dependent on their revisit behavior, we use a salon's specific idea for their customers. The salon provides 20%

TABLE 1. Summary statistics of the number of visits

Variable	**Mean**	**Med.**	**Std. Dev.**	**Min**	**Max**
# of visits	7.11	4	7.894	1	79
N			52441		

discount coupons for customers visiting 2-4 times in order to encourage regular visits and 10% discount coupons for regular customers making more than 5 visits. Also, we use the statistical information from dataset. Although the maximum visits are 79 in Table 1, median and 75 percentile are 4 and 9 times, and 90 and 95 percentile are 17 and 23 times. Therefore, in Table 2, we divide the sample into five groups according to the number of visits.

TABLE 2. Segmentations

Name	Description
non-loyal	first visit customer and visit less than two times
regular1	customers who visit 2-5 times
regular2	customers who visit 6-9 times
loyal1	customers who visit 10-19 times
loyal2	customers who visit 20 times and more

Table 3 shows the descriptive statistics for durations among loyal, regular and non-loyal customers for the haircut service. We found that the gaps for duration of revisits among the five segmentations, the mean and median grow rapidly smaller from non-loyal through loyal2. It is surprising that the loyal1 customers revisit the salon more than one month and loyal2 customers revisit almost two months earlier than non-loyal customers. Also the standard deviation and coefficient of variations are smaller from non-loyal to loyal2, meaning that there is more homogenous behavior among loyal customers about revisits than non-loyal and regular customers. Thus, we consider the differences in a customer's intensity for maintaining or changing their hairstyle in our empirical analysis.

TABLE 3. Summary statistics of duration

Variable	Mean	Median	Std. Dev.	Coef. of var.
non-loyal	112.2	76	137.8	1.23
regular1	100.6	74	111.9	1.11
regular2	89.5	70	79.4	0.89
loyal1	77.8	63	64.4	0.83
loyal2	59.8	51	42.4	0.71

3. EMPIRICAL STUDY: DURATION ANALYSIS

The above observations have an important implication in our current analysis. They suggest that the durations of visiting the salon are different for each customer depending on their frequency of visits. Thus, we could observe the different behavioral patterns between loyal and non-loyal customers. The aggregation of loyal and non-loyal customers may possibly lead to inappropriate implications of the statistical inferences. Therefore, we specify a duration model for each of them and apply the Cox promotional hazard model.

3.1. **Model.** Survival analysis is concerned with the observation of time between entry to an event and a subsequent event, e.g., death. We also find that censored survival times occur if the event of interest does not occur for an individual event during the research period. Three decades ago, the Cox proportional hazards model was well established as a statistical technique for exploring the relationship between the survival of a patient (occurrence of any interest event) and several explanatory variables. Additionally, it is the most commonly used model in hazard/intensity regression, not only in medicine and pharmacology but also several other fields, e.g. bankruptcy analysis in the economics field. In this study, applying Cox regression confers an advantage that it is able to describe the intensity function for all individuals in the sample and is in constant proportion to an unspecified baseline. Thus, without completely specifying the intensity function, we can determine the relationship between the duration of visiting the salon and the explanatory variables. With respect to this point of view, the Cox model is often called the "Semiparametric model." The Cox model, that is, the conditional intensity function given the covariate value, x and the survival function are assumed to be of the form,

$$\lambda_i(t|x) = \lambda_0(t)\exp(\beta' x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}), \tag{1}$$

and

$$S_i(t|x) = S_0(t)^{\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im})}. \tag{2}$$

where $\beta = (\beta_1, \ldots, \beta_m)$ is the vector of an unknown parameter, and $\lambda_0(t)$ denotes the baseline intensity function. From eq. (1), the influences of $x = (x_1, \ldots, x_m)$ on the intensity are in proportion to $\lambda_0(t)$ for the individual without assuming the adhoc functional forms for baseline intensity. To estimate β, the partial likelihood function is given by

$$L(\beta) = \Pi_{i=1}^{F} \frac{\exp(\beta' x_i)}{\sum_{j \in R(t_i)} (\beta' x_j)}, \tag{3}$$

where $i = 1, \ldots, F$ denotes the observed number of events. $R(t_i) = \{j : t_j \geq t_i\}$ is the number of events between time t_i and t_j. Let $\hat{\beta}$ be the maximum partial likelihood estimate of β, obtained by maximizing the partial log-likelihood function, $LL(\beta) = \log L(\beta)$. It follows that

$$LL(\beta) = \sum_{i=1}^{F} \beta' x_i - \sum_{j=1}^{F} \log \left[\sum_{j \in R(t_i)} \exp(\beta' x_j) \right]. \tag{4}$$

In fact, in order to obtain $\hat{\beta}$, we can search maximum $LL(\beta)$ using the Newton-Raphson method.

3.2. **Covariates.** We showed the construction of our target dataset - the number of customers who have used the haircut service is 11,334 within the period from 2003 to 2010 (2,418 days). In this salon, more than 90% customers are female. The mean, the median and 90 percentile of age are approximately 28, 27 and 38 years, respectively. The majority of customers at the salon are female and their ages lie between 20 and 30 years old and live within 50 km of the salon. Our dependent variables for performing Cox regression are **Duration (days)** which have been described in the previous section.

Table 4 shows the explanations of the covariates. **Age** is the customer's age on the day of their last visit, **Sex** 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. When customers make an appointment with a specific hairdresser, it is **Appointment**=1 and otherwise is 0. **Services** are dummy variables and present the received treatments for the last visit and there are three combinations that exist - e.g. 1) Cut and Color; 2) Cut and Permanent Wave; and 3) Cut, Color and Permanent Wave. The customers who only had haircuts on their last visit are base of choice in this dummy set.

In this study, we adopt three very unique pieces of information as covariates. **Salon's Congestion** presents the level of congestion in last time visits. In order to measure **Salon's Congestion**, we divide the daily number of customers by the 90 percentile of number of customers over the year. The mean is 0.75, minimum and maximum are 0.13 and 1.56, respectively. We assume the customers may pay attention to the congestion not only of the salon as a whole, but also that of their own hair hairdresser. **Hairdresser Congestion** is the daily number of customers for each hairdresser. Using the hairdresser's dummy variables, we can control the time invariant heterogeneity among hairdressers; moreover, we control the time variant skill/experiences of hairdressers by **Hairdresser's Capacity**. It is defined by the 90 percentile of number of customers in every year for the hairdressers. For these three covariates, we multiply age dummy to obtain the different coefficients by six age cohorts.

Hairdressers are the dummy variables. There are eleven hairdressers in this period. **Hairdressers** from 2 to 11 show up in the estimation result, **Hairdresser 1** has been the top hairdresser for seven years. We can summarize our covariates as three groups: [1] demographic variables (**Age**, **Sex** and **Distance**); [2] customers' selectable variables (**Appointment**, **Services** and **Hairdressers**); and

[3] customers' experiences at the salon (**Salon's Congestion**, **Hairdressers' Congestion** and **Hairdresser's Capacity**). Both of the congestion variables are unobservable and uncontrollable for the customers, and we think these congestion experiences might affect the next revisit behaviors. The hairdresser's capacity variable is proxy for each hairdresser's skill year by year. We could control unobservable time-varying heterogeneity among the hairdressers. We are interested in the influences of these variables on each customer's attitude.

TABLE 4. Covariates Descriptions

Variables Name	Descriptions
Age	age of customers on the last visit day
Sex	Dummy var., Male= 0, Female= 1
Distance	distance between residents and the salon.
Appointment	Dummy var., No appointment= 0, appointment= 1
Services	Dummy var., Last taken services, there exist three combinations, Color dummy, Perm dummy and color & perm dummy w. cut.
Salon's congestions	# of customers$_{daily}$ / 90 percentile of # of customers$_{yearly}$
Hairdresser's congestions	counted # of customers for each hairdresser, daily based data
Hairdresser's capacity	estimate 90 percentile of # of customers$_{yearly}$ for each hairdresser
age 20-45 dummy	age20 is dummy var. of $20 \leq$ age < 25 years old age25 is dummy var. of $25 \leq$ age < 30 years old age30 is dummy var. of $30 \leq$ age < 35 years old age35 is dummy var. of $35 \leq$ age < 40 years old age40 is dummy var. of $40 \leq$ age < 45 years old age45 is dummy var. of 45 and more $\leq$ age
Hairdresser 2-11	Dummy var., hairdresser in charge for the last visit

3.3. **Estimation Results.** Similar to a general multiple regression, interpreting the Cox model involves examining the coefficients for each explanatory variable. A significantly positive estimated coefficient for an explanatory variable contributes to the positive effect on intensity $\lambda_i(t|x)$, while increasing the explanatory variable leads to a higher intensity. Conversely, a significantly negative coefficient implies that the variable moves in the direction opposite to intensity. Another quality of the Cox model interpretation is the risk ratio, where $textitriskratio = \exp(\beta)$. If one assumes that $\beta = 0$ then the explanatory variable has no impact on the intensity of using a hair service. Thus, $\exp(\beta) = 1$ describes that the intensity level is equal to the baseline, or there is no impact of the explanatory variable on the intensity ($\beta = 0,$ *risk ratio=1*) in the sense of individual differences. Moreover, a *risk ratio* > 1 implies that the risk is higher than the baseline ($\beta = 0$), while the *risk ratio* < 1 implies that the risk ratio of termination is below the baseline of intensity $\lambda_0(t)$.

Table 5 provides the results of the Cox proportional hazard regression for five segments, non loyal, regular1, regular2, loyal1 and loyal2 respectively. In order to specify the functional form of each intensity function, we apply the stepwise AIC method to select the indispensable explanatory variables for the intensity functions from the 38 candidates of explanatory variables. The blank cells denote that the variables were not selected in the models as explanatory variables by the AIC criterion. The Coefficients marked $**$ and $***$ in the tables denote significantly positive or negative at the 5% level and the 1% level.

Among the five segments, the coefficients of **Sex** are significant at 5% level. Since **Sex** is coded as 0 for male and 1 for female, negative coefficients imply that male customers have a higher intensity compared to female customers. Moreover, male customers revisit the salon earlier. If we could say that male hair growth is faster than female hair growth, we could have easily explained this result. However, it is not easy for us to prove. It could be the more natural assumption for constancy of hair growth in individuals. We now attempt to carefully examine this from a different perspective - a shorter hairstyle seems to be more difficult to maintain; therefore, customers with short hairstyles will tend to increase the frequency of visits.

Generally, male customers have a shorter hairstyle than female customers; hence, this is a possible interpretation of the negative coefficient of **Sex**. Without regular1, we observed **Age** coefficients are significantly positive and the magnitude of loyal2 is the biggest. Thus, we assume that hair growth is a constant individual characteristic; then, 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. **Distance** shows 5% significant negative coefficients in non-loyal and loyal2. It implies that the customers who live further away will visit the salon later and their duration will be longer.

As a next step, we will explain about the coefficients of **Appointment**, **Services** and **Hairdressers**. These three variables are distributed as customers' choice variables. **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 intensity function. We may say that the customer who makes an appointment with their favorite hairdresser will come back to the salon sooner than the customer who does not. **Services dummy** variables have negative impact on intensity function, except loyal2's color dummy and color & perm. dummy of non-loyal and loyal2. The customers who only had a haircut on their last visit are base of choice in this dummy set. It suggests that the customers who used haircut and hair color services at the same time for their last visit would have longer durations of non-visit than those customers who had had haircuts. Likewise, the coefficient of haircut and permanent wave and haircut, coloring and permanent wave are also negative, because it implies that hair color and permanent wave services would delay a customer's revisit. We adopt **Hairdressers** to control time-invariant heterogeneity effects among the eleven hairdressers. The base hairdresser is a **Hairdresser1** and he has been a top hairdresser for seven years. Both loyal1 and loyal2 have positive coefficients among hairdressers other than Hairdresser1, and other non-loyal segments have significant negative coefficients. We could observe the difference in the way hairdressers affect the revisit behavior in the non-loyal and loyal group.

Finally, we intend to consider the effects of congestion and hairdresser's skill on revisit behavior. We assume that the customers who visit the salon have some impressions or experiences about the salon's atmosphere, congestion and skills in addition to undergoing treatment. In this study, we adopt the salon's congestion, the hairdresser's congestion and the capacity. Especially, the capacity variable is a proxy for hairdresser skill. If their coefficients are significantly positive, they increase the intensity, that is, the customers will visit the salon sooner (the duration will be shorter). In a negative scenario, we can explain the reverse. Table 6 shows the summary of the results about these three variables among the segments. We are interested in the effects of the new variables on the next revisit probability and also if different impacts on the intensity exist within/between segments and age cohort. 0 denotes that the coefficients are insignificant, – and + are significantly negative or positive at 5% level. $*$ means the coefficients' magnitude is significantly different from others. We expected the hairdresser's skill and experiences to have positive effects on the next revisits. In the customers' results of loyal1 and loyal2, we could find **Hairdresser's Capacity** is significantly positive. On the other hand, we observed the negative coefficients in the other three segments. In the non-loyal and loyal2 segments, **Hairdresser's Congestion** is significant for all age cohorts; interestingly, among the non-loyal customers, the impact of the hairdresser's congestion is homogeneously negative. We found that the different response to the hairdresser's congestion was dependent on the age group in loyal2 segment.

When we are in a busy salon, two different aspects may exist; one is that the salon's impression is active and another is that it is impulsive. We expect that if the customer prefers the active salon, the coefficient could be positive, and if they feel rushed, the coefficient might be negative. We focus on the results of loyal customers, who visit the salon more than twenty times (loyal2); there was a clear difference between under-30s and over-30s. **Salon's Congestion** can increase the intensity of their next visit, however congestion delays the next revisits for those over 30 years old. We can also find the negative effects of congestion on the intensity function in other segments, and these are all negative coefficients.

In our view, regular1 and 2 seem to be similar results, and we might combine these two in the same segment. Remarkably, the result of the loyal2 segment shows the necessity to include these variables for the intensity function. These new adopted variables influence the customer's attitude toward the duration of visiting the salon. It suggests that loyal customers are more sensitive about the salon's atmosphere, congestion, hairdressers' skills and their own experiences at the salon than less loyal customers and their last visit experiences would lead to differences in the non-visit interval among the customers. This subsection summarizes that intensity functions of the five segments have been specified by different models.

Table 5: Results of Cox Regression

Segment	non-loyal $n = 1$	regular$_1$ $2 \leq n \leq 5$	regular$_2$ $6 \leq n \leq 9$	loyal$_1$ $10 \leq n \leq 19$	loyal$_2$ $20 \leq n$
	$\hat{\beta}$ /se	$\hat{\beta}$/se	$\hat{\beta}$/se	$\hat{\beta}$/se	$\hat{\beta}$/se
age	0.019***		0.008**	0.008**	0.028***
	(0.003)		(0.003)	(0.003)	(0.004)
sex	-0.120**	-0.244***	-0.266***	-0.348***	-0.478***
	(0.042)	(0.026)	(0.035)	(0.033)	(0.044)
distance	-0.004**	-0.002	-0.002	-0.002	-0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
appointment	0.216***	0.372***	0.332***	0.344***	0.457**
	(0.027)	(0.026)	(0.049)	(0.059)	(0.141)
color dummy	-0.092**	0.029	-0.128***	-0.074**	0.097**
	(0.029)	(0.019)	(0.027)	(0.028)	(0.041)
perm dummy	-0.182***	-0.067**	-0.188***	-0.091**	-0.131**

Table 5: Results of Cox Regression

Segment	non-loyal $n=1$	regular$_1$ $2 \leq n \leq 5$	regular$_2$ $6 \leq n \leq 9$	loyal$_1$ $10 \leq n \leq 19$	loyal$_2$ $20 \leq n$
	$\hat{\beta}$ /se	$\hat{\beta}$/se	$\hat{\beta}$/se	$\hat{\beta}$/se	$\hat{\beta}$/se
	(0.034)	(0.024)	(0.035)	(0.036)	(0.053)
color & perm		-0.099**	-0.295***	-0.191***	
dummy		(0.036)	(0.051)	(0.053)	
salon's					2.944***
congestions					(0.675)
Hairdresser's	-0.015***				-0.182**
congestions	(0.004)				(0.059)
Hairdresser's	-0.051***	-0.027***	-0.014**	0.041***	0.065***
capacity	(0.009)	(0.006)	(0.005)	(0.006)	(0.013)
salon's cong.		-0.116			-2.812***
* age20 dummy		(0.050)			(0.729)
salon's cong.				-0.230**	-2.761***
* age25 dummy				(0.075)	(0.680)
salon's cong.					-3.092***
* age30 dummy					(0.690)
salon's cong.			-0.237**		-2.957***
* age35 dummy			(0.081)		(0.701)
salon's cong.			0.415		-3.223***
* age40 dummy			(0.227)		(0.713)
salon's cong.	-0.684***		-0.540**		-2.996***
* age45 dummy	(0.136)		(0.230)		(0.706)
Hairdresser's cong.		0.011**	0.012**	0.013	0.198**
* age20 dummy		(0.004)	(0.006)	(0.007)	(0.063)
Hairdresser's cong.			0.008	0.010	0.174**
* age25 dummy			(0.005)	(0.006)	(0.060)

Table 5: Results of Cox Regression

Segment	non-loyal $n = 1$	regular_1 $2 \leq n \leq 5$	regular_2 $6 \leq n \leq 9$	loyal_1 $10 \leq n \leq 19$	loyal_2 $20 \leq n$
	$\hat{\beta}$ /se	$\hat{\beta}$/se	$\hat{\beta}$/se	$\hat{\beta}$/se	$\hat{\beta}$/se
Hairdresser's cong.	0.021				0.185**
* age30 dummy	(0.010)				(0.060)
Hairdresser's cong.		0.006		0.020	0.162**
* age35 dummy		(0.003)		(0.010)	(0.061)
Hairdresser's cong.		-0.029			0.201**
* age40 dummy		(0.015)			(0.063)
Hairdresser's cong.			0.029	0.020	0.179**
* age45 dummy			(0.015)	(0.013)	(0.060)
capacity	0.006**		-0.013	-0.020**	
* age20 dummy	(0.002)		(0.006)	(0.007)	
capacity			-0.014		0.034**
* age25 dummy			(0.005)		(0.012)
capacity	-0.018	-0.005**	-0.017***	-0.013***	0.033**
* age30 dummy	(0.008)	(0.002)	(0.004)	(0.003)	(0.013)
capacity	-0.008			-0.030***	0.038**
* age35 dummy	(0.005)			(0.009)	(0.015)
capacity		0.028	-0.035		0.033
* age40 dummy		(0.012)	(0.014)		(0.017)
capacity				-0.018	
* age45 dummy				(0.011)	
Hairdresser2	0.141***		-0.091	-0.315***	
	(0.039)		(0.044)	(0.077)	
Hairdresser3	-0.376***	-0.177***	-0.132***	-0.297***	
	(0.060)	(0.036)	(0.039)	(0.056)	
Hairdresser4	-0.852***	-0.244***	-0.382***		

Table 5: Results of Cox Regression

Segment	non-loyal $n=1$	regular$_1$ $2 \leq n \leq 5$	regular$_2$ $6 \leq n \leq 9$	loyal$_1$ $10 \leq n \leq 19$	loyal$_2$ $20 \leq n$
	$\hat{\beta}$ /se	$\hat{\beta}$/se	$\hat{\beta}$/se	$\hat{\beta}$/se	$\hat{\beta}$/se
	(0.107)	(0.066)	(0.065)		
Hairdresser5	-0.652***	-0.177**	-0.111	0.336***	0.603***
	(0.098)	(0.060)	(0.044)	(0.054)	(0.126)
Hairdresser6	-0.534***	-0.161***		0.176***	0.327***
	(0.068)	(0.040)		(0.032)	(0.043)
Hairdresser7	-0.800***	-0.145**		0.275***	0.676***
	(0.090)	(0.054)		(0.048)	(0.084)
Hairdresser8	-1.199***	-0.299***	-0.250**		1.292***
	(0.139)	(0.088)	(0.089)		(0.164)
Hairdresser9	-0.958***	-0.222**	-0.292***	0.291***	0.907***
	(0.112)	(0.071)	(0.067)	(0.076)	(0.134)
Hairdresser10	-0.965***	-0.207**		0.373***	0.718***
	(0.125)	(0.074)		(0.082)	(0.147)
Hairdresser11	-0.573	-0.639	-0.675		1.234***
	(1.006)	(0.361)	(0.381)		(0.322)
AIC	113794.23	278142.47	131434.38	129000.37	50403.91
# of Obs	11246	18589	9169	8894	3867

TABLE 6. Estimation results of Congestions of salon and Hairdressers and Hairdressers' capacity (Summary)

	non-loyal			regular1			regular2			loyal1			loyal2		
age	Salon	Sylist	Capa.	Salon	Sylist	Capa.	Salon	Sylist	Capa.	Salon	Sylist	Capa.	Salon	Sylist	Capa.
10-19 (base)	0	–	–	0	0	–	0	0	–	0	0	+	+	–	+
20-24	0	–	–*	0	+*	–	0	+*	–	0	0	+	+*	+*	+
25-29	0	–	–	0	0	–	0	0	–	–*	0	+*	+*	–*	+*
30-34	0	–	–	0	0	–*	0	0	–*	0	0	+	–*	+*	+*
35-39	0	–	–	0	0	–	–*	0	–	0	0	+*	–*	–*	+*
40-44	0	–	–	0	0	–	0	0	–	0	0	+	–*	+*	+
45 and more	–*	–	–	0	0	–	–*	0	–	0	0	+	–*	–*	+

0 denotes the coefficients are insignificant, – and + are significantly negative or positive at 5% level.

$*$ means the coefficients' magnitude are significantly different from others.

3.4. **Comparative Revisits Rates among the segments and the prediction method.** In this subsection, we apply the Cox model estimation results to calculate the aggregate revisit rates for each segment and show how to predict each customer's revisit rates and examine the standard error to the interval estimation. We obtain $\lambda_0(t)$ estimator by the Kaplan-Meier method. We then estimate the cumulative baseline intensity function $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$, and consider the differences of individual intensity. Then the baseline survival (Revisiting) function of Cox regression can be written as,

$$\hat{S}_0(t) = \exp\{-\hat{\Lambda}_0(t_i)\}q_i, \text{ where } q_i = -\frac{\exp(\hat{\beta' x_i})}{\sum_i \sum_j \exp(\hat{\beta' x})}^{\frac{1}{\exp(\hat{\beta}' x_i)}}.$$

We can rewrite $\hat{S}_0(t) = \hat{S}_0(t_i)q_i$, when $\beta = 0$, q_i equals to 1; this estimator results in the ordinary Kaplan-Meier curve $(\hat{S}_0(t_i))$. Moreover, in the Cox proportional intensity model, the survival function $S(t|x)$ of an individual with covariate values x is given by $S(t|x) = S_0(t)^{\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im})}$. The estimate $\hat{S}(t|x)$ is a right-continuous step function, with jumps in the event times. This indicates each customer's probability of revisiting and its duration. It is referred to as the Adjusted KM curve. Figure 1 yields the baseline revisiting function for the non-loyal, regular and loyal customers, respectively. The horizontal axis indicates the duration (in days) and the vertical axis indicates the probability of revisiting the salon. In Figure 1, for the median (50%) revisiting rate among the five segments, the interval within which no visits are made is 187 days, 85 days, 77 days, 69 days and 54 days for non-loyal, regular1, regular2, loyal1 and loyal2 customers. We can observe the different shapes of baseline intensity functions among the five segments.

To predict the individual revisit day, we define the non-revisit rate function by Cox regression results as below. We abstract customers who visited last time within one month from the end point which is explained below.

$$S_i(\hat{t}|x)_{Cox} = S_0\hat{(t)}^{\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im})} \tag{5}$$

With selected samples visiting within one month l, we consider the binomial distribution for their revisit behavior. When they revisit the salon equals 1 and

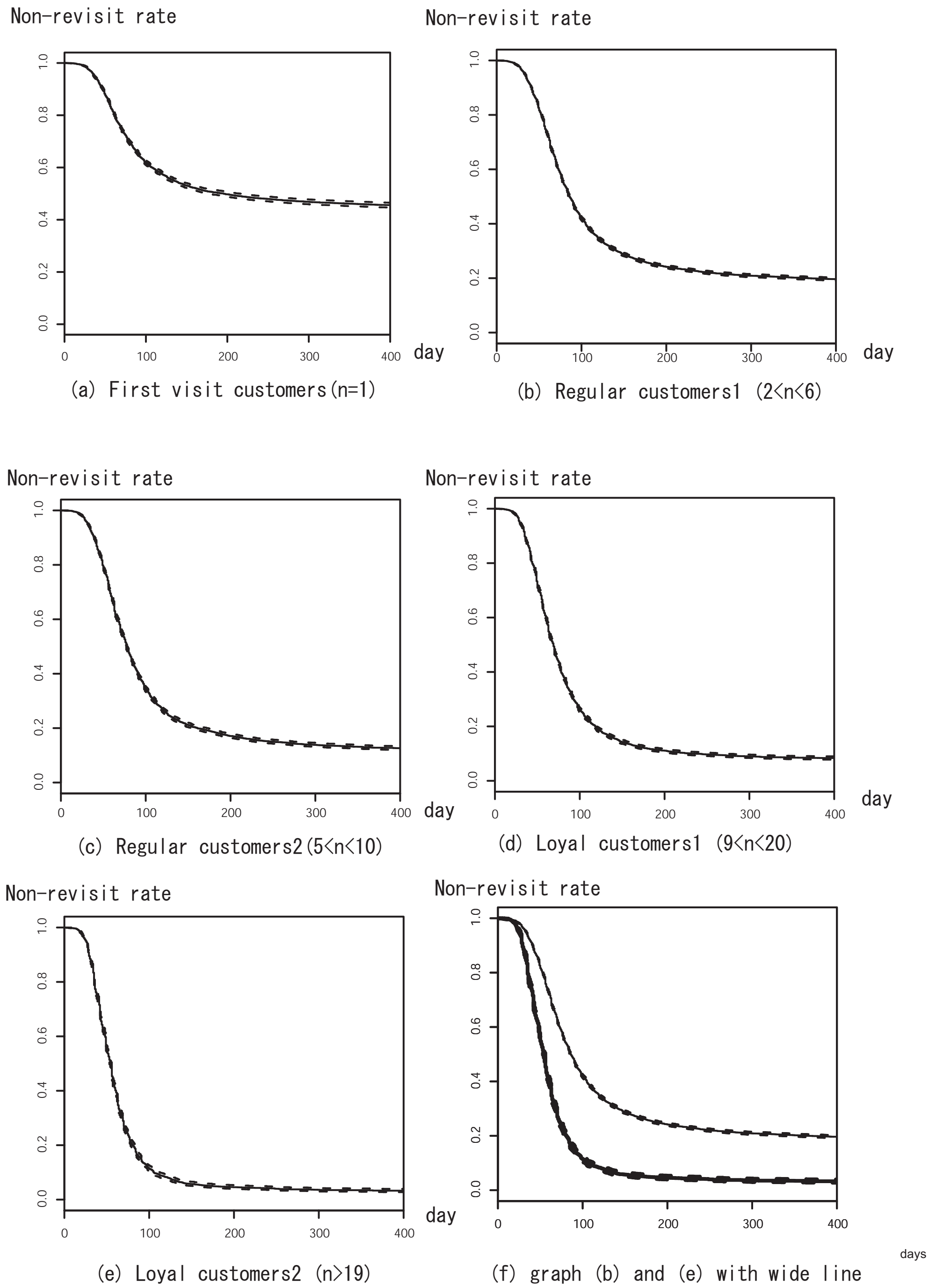

FIGURE 1. Estimated Non Revisit Rates (Base line intensity function)

otherwise are 0 as eq. (6).

$$E\{1_{\{T_i \leq t_i\}}\} = P\{1_{\{T_i \leq t_i\}}\} = 1 - S(t_i|x) \tag{6}$$

In order to calculate the revisit rates for each customer within any days (e.g. 30 days or 100 days) since the last visit by eq. (6), we could obtain the revisit probability for the hair salon when $t > 30$ or $t > 100$ using eq. (7). $\hat{S}(t)$ is Normal distribution with mean $S(t)$ and standard deviation $SE(t)$ asymptotically. We obtain the standard error and the 95% confidence intervals in eq. (8).

$$(7) \qquad Revisitrates = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \hat{S}_0(t_i)^{\exp(\hat{\beta}' x_i)}\right)$$

$$(8) \qquad StandardError = \frac{1}{\sqrt{n}}\sqrt{\left(1 - \hat{S}_0(t_i)^{\exp(\hat{\beta}' x_i)}\right)\left(\hat{S}_0(t_i)^{\exp(\hat{\beta}' x_i)}\right)}$$

In order to know when the customers will revisit the salon, we estimate the median revisit duration for each of them using $\hat{S}(t|\mathrm{x})_{cox}$ and obtain the prediction value of their next visit date for the salon from the spontaneous point. The prediction of revisit rates tell us when and how many customers will visit a salon and which services they will receive.

4. Concluding Remarks

In this study, we apply duration analysis to specify a model of a hair salon customers' behavior regarding their visits and then estimate their aggregated revisit rates/dates. We 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.

The dataset of this study was obtained from a hair salon located in Japan's third largest city, Osaka, and our target customers primarily used the haircut service. We adopt the following approach: [1] assuming that there are differences among first visit customers, regular and loyal customers with respect to the intensity function: [2] introducing customer's behavioral variables, the hair salon's congestion and hairdresser's skill variables in addition to demographic variables; [3] observing the degree of sensitivity to the congestion and skill to intensity function by age cohorts; [4] by applying the estimation results of Cox regression, we examine the aggregated revisit rate and show the procedure of measuring the individual next revisit date.

As a result, we found that the intensity functions of non-loyal and loyal customers have been specified by different models. For loyal customers, when they had higher skilled hairdressers on their last visit, they tend to come back to the salon sooner. We could also observe differences between the first visit customers and loyal customers in terms of the responses against hair salon and hairdresser's congestion. The hairdresser's congestion has a negative effect on their revisit rate. In the loyal customers' results, younger customers prefer the hair salon's active atmosphere, but customers over the age of 30 will revisit much later if they had a treatment on a busy day. It suggests that customer heterogeneity should be included in the intensity model and we also need to consider the information about the hair salon (supply side).

References

1. Akaah, I. P., Korgaonkar, P. K. and Lund, D. (1995): Direct marketing attitudes, *Journal of Business Research* **34**, 211–219.
2. Cox, D. R. (1972): Regression models and life tables (with discussion)., *Journal of the Royal Statistical Society,* **Series B 34**, 187– 220.
3. Huges, A. M. (1994): Strategic Database Marketing: The Master Plan for Starting and Managing a Profitable, Customer- Based Marketing program., *Probus, Pub Co.*
4. Jonker, J. J., Piersma, N. and Van den Poel, D. (2004): Joint Optimization of customer segmentation and Marketing policy to Maximize Long- term Profitability *Expert Systems with Applications* **27**, 159–168.
5. Korgaonkar,P. K.,Lund, D.and Price, B. (1985): A StructuralEquations Approach Toward Examnination of Store Attitude and Store Patronage Behavior., *Journal of Retailing,* **61**, 39-60.
6. Vigna, D. S and Malmendier, U. (2006): "Paying Not to Go to the Gym," *The American Economic Review*, Vol.96, No.3, pp. 694-719

Current address: Cowles Foundation Yale University, 30 Hillhouse Avenue New Haven CT 06511, U.S.A, Research Institute of Economy, Trade and Industry, Tokyo 100-8901, Japan.

E-mail address: yoko-konishi@rieti.go.jp