An Empirical Study of Customer's Repeat Visit Frequency on the Internet

Suke Kyu Lee*

〈Abstract〉

This study explores whether a NBD type of model can be applied to characterize the underlying frequency distribution of online consumer's visit behavior. In this study, the following two research questions are addressed: (1) How can we characterize the underlying distribution pattern(s) of the number of repeat visits to a site? (2) How can consumer's Internet usages and his/her demographics affect the average number of visits to the site?

Through the empirical investigation, this study found that NBD models are directly applicable to characterize the underlying distribution of visit frequency on the Internet. Furthermore, this study addresses some managerial implications for understanding how site visits are determined. Especially this study highlights the relationship between repeated visits and the visitors' Internet Usages and demographics. The proposed models are estimated and validated by online panel data that covers more than 1000 different sites and has 800,000 observations.

Key Words : internet browsing behavior, NBD model, web panel data analysis, site loyalty

Ⅰ. Introduction

The information on the expected number of visits to a site is important to business managers. Imagine first that you are a marketing manager in Yahoo!. You can review the list of existing customers who have already made visits and who have registered as members on Yahoo!. One of main revenue sources of Yahoo! comes from hosting Web advertising on its site. Many companies buy the advertising spaces, in the form of banner advertisements, available at Yahoo!. The rates of Web advertising depend on Web browsing behaviors of visitors to the site where companies want to put their advertisement. Business managers need to know the 'visit frequency' as well as the 'visit timing' of each customer.

Once a manager understands and predicts customers' browsing behaviors, s/he can charge higher advertising rates, since companies are willing to pay more money to reach the right customers. As online panel data is available, it is easy for the managers to get more information on the complete histories of each customer's Internet use. Using this panel data,

* Assistant Professor of Marketing at Hankuk University of Foreign Studies' School of Business Administration.

business managers can figure out how many customers make visits during a given period of time. Furthermore, managers can predict how many customers will visit again and how many new visitors will come to their sites at a given period of time. Thus, the analysis of visit frequency is helpful in planning marketing strategies for advertising and promotion.

In addition, consumer repeat visit behavior on the Internet is unique in terms of the nature of the data itself. That is, the repeat visit dataset is a typical case of "Ultra-High-Frequency Data,' a recent hot issue in econometrics. According to Engle (2000), the 'ultra-high frequency data' has one salient and critical feature in that they are fundamentally irregularly spaced. This irregularity causes very complex econometric problems. Kim and Park (1997) addressed the importance of investigation on the irregularity by incorporating each individual's heterogeneity into a model of shopping trip patterns.

Despite the importance of understanding consumer visit frequency on the Internet and the uniqueness of online panel data in terms of 'irregularity and ultra high frequency,' no study has yet investigated online consumer visit behavior explicitly. Thus, this study examines the following two questions:

1) How can we characterize and model the underlying distribution pattern(s) of the number of repeat visits to a site?
2) How can a consumer's Internet use and his/her demographics affect the average number of visits to the particular site?

To answer the questions, this study first investigates the empirical pattern of visit frequency to identify the underlying stochastic distribution and the dynamics of the distribution over time. Through the empirical investigation, this study addresses the direct applicability of either NBD or CNBD models to characterize the underlying distribution of visit frequency on the Internet - since the repeat visit behavior can be considered a 'frequently occurring behavior.' In marketing, the 'frequently behaviors' have been successfully investigated in various contexts by using NBD types of models.

First, many models of both 'brand choice' and 'purchase timing' have been formulated using the negative binomial distribution (NBD) model frameworks (Ehrenberg 1959, 1972; Gupta 1991; Herniter 1971; Jeuland, Bass and Wright 1980; Morrison and Schmittlein 1981; Zufryden 1977, 1978). The choice models based on product or brand-purchasing behavior have addressed the frequency issue in the context of brand or product purchases. The common applications of the NBD model framework were also

successfully extended to both the 'Filler Trips to Grocery Stores' (Frisbie 1980) and the 'Shopping Mall Visits'(Dunn, Reader and Wrigley 1983).

Thus, it should be interesting to investigate whether a NBD type of model framework, widely used in previous marketing models, can also explain online consumer browsing behavior. The reason is that online consumer visit to a site can be considered a kind of ultra-high frequency data, making the online data unique, when compared to the previous 'frequently occurring behavior (brand/ product purchases, filler trips to grocery stores, and shopping mall visits).

Further more, this study contrasts a NBD model with a CNBD model in terms of predictive accuracy. Based on the comparison, this study describes the effect of covariates on the visit frequency by using a Negative Binomial (NB) Regression. The NB regression is a generalization of the Poisson regression (Allison 1999). Through the NB regression analysis, this study illuminates the covariate effects and the dynamics of visit frequency over time. Particularly, the study addresses how the average number of visits to a particular site during a given period of time is affected by consumer Internet use and her/his demographic variables.

This study makes the following contributions. From a theoretical viewpoint, this study first specifies a pattern of visit frequency distribution on the Internet and then gives a solution to the question of what type of statistical distribution can characterize the underlying distribution pattern well. This study concludes that a CNBD model or a NBD model can be directly applied to characterize the underlying distribution. Thus, this study provides another application of a NBD model framework (widely used in purchase behavior models) to explain online consumer's visit behavior. By characterizing 'visit to a site' as a "repeatedly occurring behavior," this study argues that consumer visit frequency on the Internet is also explained by a NBD type of model such as those which have been successfully applied to 'filler trips to grocery stores,' or 'repeat visits to a shopping mall.' Thus, given the similarity of visit frequency on the Internet, this study provides another natural extension of NBD type of models to modeling online consumer behavior.

From a practical viewpoint, this study provides managers with information on how many visits are made either (or both) by new customers or (and) by existing customers during a given period of time. By using the proposed model, managers can also predict how many customers will return and how many new visitors will come to a site at a given

period of time. The prediction of when consumers will come back is critical to develop advertising and promotional strategies. Second, after comparing the predictive accuracy of the proposed CNBD model and the base model (NBD), the study found that both models provide reasonably good estimates of the actual conditional probability values. But there are only slight discernable differences between the estimates of the two models. Thus by using a NBD model instead of a CNBD model, managers can evaluate their marketing activities by comparing the distribution patterns before and after any efforts are made.

The study developments are organized as follows: First, we describe the model development for visit frequency. In the model development, we start with a NBD model as a base model and then investigate a CNBD model to characterize the underlying distribution of visit frequency. In each model, the main model assumptions, specifications, estimation procedures are discussed. Second, we provide an empirical illustration to estimate and validate the proposed models by using online panel data and then contrast the two models. Third, based on the empirical results, we investigate the effect of consumer Internet use and his/her demographic on the average number of visits by using a Negative Binomial Regression. Fourth, we address the managerial implications of the model of visit frequency. Finally we summarize the contributions and limitations of the study, and suggest research opportunities for future study.

Ⅱ. Model Development of Visit Frequency

1. A NBD Model

1) Model Assumptions

This study uses a NBD model as a base model to characterize the underlying distribution of visit frequency on the Internet. NBD models assume that the individual visiting rates follow a Poisson distribution (DeGroot 1984). Thus, a Poisson process with an average visit rate l generates each consumer's visit to a site. In a given period of time, the resulting Poisson distribution for the number of visits to a Web site is easy to handle.

Although many researchers have expressed concerns with the memoryless property of the exponential distribution (Chatfield and Goodhardt 1973), the assumption has been widely used as a starting point or a base model to explain 'frequently repeated' consumer behavior (Gupta 1991; Schmittlein, Albert, and

Morrison 1985; Schmittlein and Morrison 1983; Schmittlein, Morrison and Colombo 1987). The behavioral aspect can be also justified on the Internet by the following reasons. First, it makes sense that each consumer has the same level of visit occurrence during a given period of time because there is no 'stockpiling effect' (inventory effect in brand purchase studies) in browsing behavior. Consumers who have just visited a certain search site (i.e., Yahoo!) are equally likely to visit the site again compared to consumers who have not visited the site for a long time. Second, the actual mode of visit frequency distribution supports the exponential assumption because the mode of the exponential distribution is 0. Despite the concerns, due to the simplicity of the exponential assumption, this study begins with the Poisson distribution of visit rates.

The second assumption is that visiting rates λ are distributed according to a gamma distribution across the population of consumers (Gamma Heterogeneity). This gamma mixing distribution on λ combines easily with the individual-level Poisson and exponential distributions. The resulting NBD appeals model foundation for many useful predictions about purchase patterns in marketing (Ehrenberg 1959, 1972; Gupta 1991; Jones and Zufryden 1980; Morrison and Schmittlein 1988). Thus, this paper also employs a gamma mixing procedure to capture the heterogeneity across consumers.

2) Model Specifications

An observable probability mixture distribution with the name of NBD results from an unobservable gamma mixing with the purchasing rates for individual-level Poisson repeats visits. Mathematically, the mixing procedure is as follows (Morrison and Schmittlein 1988):

$$Pr_{NBD}(X=k \mid a, q) = \int^{[0,\ x]} Pr_{poisson}(X=k|\lambda)\ f(\lambda|\ a,b)\ d\lambda$$

where the specific functional forms are:

$$Pr_{poisson}(k \mid \lambda, T) = Exp(-\lambda T)[\lambda T]^k / k!, \quad \text{for } k=0, 1, 2, \cdots\text{.., etc.}$$

$$f(\lambda|\ a,b) = b^a \exp(-b\lambda)^{a-1} / \Gamma(a)$$

The resulting distribution is the NBD model. That is, we get:

$$Pr_{NBD}(k \mid T) = \{\Gamma(a+k)/\Gamma(a)k!\}\ q^k (1-q)^a$$

with parameters (a and q) where $q = T/(T+b)$ and $E(k) = E(\lambda) = aT/b$, $Var(k) = aT/b + aT/b^2 = E(\lambda) + Var(\lambda)$. The variance of the observable NBD mixture has two components: the average within-individual Poisson variation (i.e., aT/b), since the variance of the Poisson is

equal to its mean) plus the across-individual variability of repeat visit rates (i.e., $var(\lambda) = aT/b^2$).

3) Estimating the NBD Parameters: a and b

This study uses "Mean-Zero Method" to estimate the model parameters. According to Morrison and Schmittlein (1988), we estimated the parameters as follows: First, we obtain observed values for proportion of consumers making 0 visits to the site (P(0)) and then counts the average number of visits per consumer (m). Second, we find a such that $[a/(m+a)]^a - P(0)=0$. The parameter b in the gamma function can be calculated from the mean formula as b = aT/m. Once we obtain the two parameters (a and b), we generate the NBD probability that consumers will make k visits to the Web site of interest during a certain period of time (i.e., 4 weeks for this study) by computing $P(0|T)= (1-q)^a$, for $k \geq 0$ and then $P(k|T)=P(k-1|T)q(a+k-1)/k$, for $k \geq 1$.

4) Conditional Expectations

Perhaps the most managerially relevant construct that results from this model is the conditional expectation E[X2 | X1=k]. This is the expected number of repeat visits made in a future period given that a consumer made k visits during the period of observation. Morrison and Schmittlein (1988) derived these conditional expectations. They argued that the conditional expectations are a linear function of k. That is, $E[X2 \mid X1 = k] = E[\lambda \mid X1 = k]$. Specifically when the mixing distribution on the individual repeat visit rates l is gamma with a shape parameter a and a scale parameter b (in this parameterization E(l) = a/b), the conditional expectations are:

$$E[X2 \mid X1 = k] = a/(b+1) + [1/(b+1)]*k, \quad k=0, 1, 2, 3, \cdots etc.$$

This equation assumes that the observation period is of unit length and the future period is also the same unit length of time. If the future time is a factor T of the observation time then the right hand side of the equation is multiplied by T.

Thus, this study utilizes the conditional expectation concept to investigate the important managerial question of how to forecast cumulative Web site penetration as well as the issue of trial and repeat purchase measures. These concepts are defined as follows:

New Triers: Pr(k≥1 in T2 given k=0 in T1)

Repeat Visitors: Pr(k≥1 in T1 and k≥1 in T2)

Repeat Visit Ratio: Pr(k≥1 T2 given k≥1 in T1)

2. A CNBD Model of Visit Frequency

1) A NBD Model

(1) Model Assumptions

The Poisson assumption may not be appropriate when there is a regularity in consumer browsing behavior on the Internet. According to Schmittlein and Morrison (1983), the purchase frequency distribution is explained well by a 'condensed' Poisson distribution (CPD). The condensed Poisson is a distribution that can be applied to events that are more regular than the simple Poisson. Many studies have used the CNBD assumption to explain purchasing behavior (Chatfield and Goodhardt 1973; Morrison and Schmittlein 1981, 1983; Zufryden 1978). Thus, this study explores whether a CNBD model is a good alternative to the NBD model to explain the underlying distribution of the browsing behavior on the Internet.

(2) Model Specification

To take into account any possible non-monotonic shape of the distribution of visit frequency (the non-monotonic shape means a regularity in the browsing behavior on the Internet), this study formulates the inter-visit time distribution as the Erlang family of density functions. The Erlang process has been used to describe the underlying stochastic process in the purchasing behavior (Herniter 1971; Jeuland, Bass, and Wright 1980; Wu and Chen 1999). Following Herniter (1971), we model consumer visit behavior on the Internet as an Erlang distribution with parameters (λ, γ):

$$f(t / \lambda, \gamma) = [(\gamma\lambda)^{\gamma} t^{\gamma-1} e^{-\gamma\lambda t}]/[\gamma-1]!, \quad \gamma, \; t \geq 0, \; \gamma=1, 2, 3, \cdots..$$

The model is a function of λ and γ, where λ is the inverse of the expected value of t (i.e., $\lambda = 1 / E(t \mid \lambda, \gamma)$, and γ is the order of the Erlang. Note that the order of Erlang (γ) should be an integer and that the γth order Erlang is the convolution of γ exponentials, all of which have the same mean value $\lambda/(\gamma\lambda)$. When $\gamma = \lambda$, the Erlang is reduced to the exponential density function, which is the base model in this study. But the main purpose of this study is not to develop a comprehensive framework, but to examine the possibility of direct application of a NBD type of model to the explanation of consumer visit behavior on the Internet. Thus, the study covers the 2nd order Erlang distribution.

Following previous studies (Chatfield and Goodhardt 1973; Gupta 1991; Herniter 1971; Kim and Park 1997; Morrison and Schmittlein 1983; Wheat and Morrison 1990; Wu and Chen 1999), we propose the Erlang-2 as a distribution of inter-visit times. Thus, we can get the following density function, survival function, and the

corresponding hazard function:

$$f_i(t) = \lambda_i^2 t e^{(-\lambda_i t)}$$

$$S_i(t) = (1+\lambda_i t) e^{(-\lambda_i t)}$$

$$h_i(t) = \frac{f_i(t)}{S_i(t)} = \frac{(\lambda_i)^2}{(1+\lambda_i t)}$$

Some customers visit a particular Web site more frequently than others. Thus, this study incorporates the heterogeneity in visit frequency rates into the model by allowing λ to have a Gamma distribution with parameters (a, b). The corresponding density function of λ is:

$$g(\lambda|a,b) = \frac{b^a e^{-b\lambda} \lambda^{a-1}}{\Gamma(a)}$$

where a, b = constant parameters (a, b > 0), and E(λ) = a / b and Var(λ) = a / b^2.

(3) Parameter Estimation of the CNBD Model

Since a CNBD model is a special case of NBD models (a CNBD assumes that inter-event times between every other events are exponential), with some restrictions the CNBD model can also be estimated by the "Mean and Zero Method" applied to NBD models. This study also follows the estimation procedure suggested by Chatfield and Goodhardt (1973) and elaborated by Zufryden (1977). The CNBD distribution parameters are estimated by equating the observed and theoretical values of the mean and the proportion of zeros.

What is the expected probability E[v(k | T)] that k visits will make during an arbitrary period of time (0,T)? In the given time-period, the distribution of visits in the whole population is obtained by mixing (or compounding) the condensed Poisson distribution with the gamma distribution (Chatfield ad Goodhardt 1973). Thus, the expected probability of k visits by a customer is computed by integrating the probability of a customer making k visits within a period of time (0,T) over the population distribution of λ. Zufryden (1977) calculated E[v(k | T)] as follows:

$$E[v(k/T)] = \int_0^\infty v(k/\lambda, T) g(\lambda/a, b) d_\lambda$$

, k=0, 1,2,

After deriving the parameters of the CNBD model, Zufryden (1977) noted that the CNBD can be stated as:

E[Pr(k visits)] = E[v(k/T)] = PCN (k)
= PN (0) + (1/2) PN (1), for k=0,
= (1/2) PN (2k-1) + PN (2k) + (1/2) PN (2k+1), for k=1, 2, 3, ….
where,

$$\text{PN (k)} = \frac{\Gamma(k+a)}{\Gamma(a)k!} Z^k (1-Z)^a$$

, k= 0, 1. 2, 3, ….

is the NBD and Z= 2T/(2T+b). By

applying this formula to online panel data, in the empirical illustration we estimate the model to characterize the expected number of visits to a particular site during a given period of time (i.e., 4 weeks).

3. A NB Regression Model

1) Model Specificationl

From the empirical comparison of a NBD model with a CNBD, this study found that there is no significant difference between the two models. However the NBD model was shown to be slightly better than the CNBD in terms of c2 values (For details, please refer to Table 1). This result is consistent with other previous studies (Morrison and Schmittlein 1988; Zufryden 1977, 1991). A CNBD model is more mathematically complicated than a NBD model. If both models provide a same level of accuracy, there is no reason to use the more complicated one.

Thus, according to 'the Rule of Parsimony' of model selection, this study adopts a NBD type of model. Specifically, this study uses a 'Negative Binomial (NB) Regression' (based on the exponential assumption of inter-visit time and gamma heterogeneity of the average number of visits) to examine the effects of consumer Internet use and his/her demographics on the distribution of average visit frequency, λ during a given period of time. The NB regression is simple and can be easily estimated by using statistical software commercially available now (i.e., SAS or STATA).

Let's assume first that the dependent variable (the number of visits k) has a Poisson distribution with expected value λi conditional on the error term (εi). We further assume that exp (σei) has a standard gamma distribution (Allison 1999). The unconditional distribution of the number of visits is a Negative Binomial distribution. Thus, we specify the NB regression model as follows:

$$Log\ \lambda_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} \ldots . + \beta_k x_{ik} + \sigma\varepsilon_i,$$

where λi is the expected value (mean) of the number of visits by consumer i and x's are explanatory variables.

2) Estimation Procedure

The Negative Binomial regression model can be efficiently estimated by a Maximum Likelihood method. We use a SAS macro program to estimate the log NB regression model (Hilbe 1994). The estimation results are summarized in Table 2 which is given in the following empirical illustration.

Ⅲ. An Empirical Illustration

1. Data

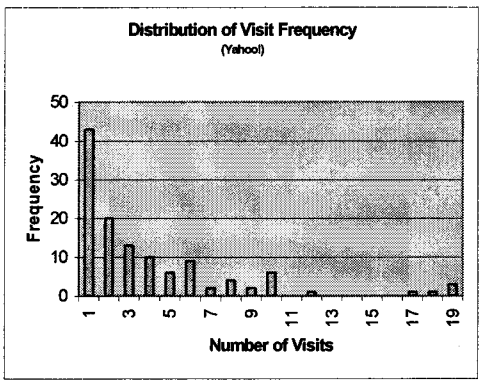
The online panel data used in this study has a total of 88,127 visit occasions of 1,529 different households in the United State. The data covers more than 1,000 different Web sites and has been recorded from the sample with size 30,000 panel members. To make it simple, we selected one portal site (Yahoo!). The empirical investigation of Yahoo! gives us a sufficient insight to estimate and to validate the proposed model of visit frequency. In addition, Yahoo! satisfies the basic conditions required by the proposed models including 'high levels of traffic' and 'relative importance' of the site on the Internet. However, the methodology used in the study to empirically evaluate Yahoo!, could be applied to any other Web site with reasonably high traffic.

2. Model Estimation

1) NBD Model Estimation

Using online panel data, this study empirically investigates the distribution of visit frequency to Yahoo!. The actual distributions found in the study are given in <Figure 1>. In addition to the justifications for the NBD assumption explained before, the actual distributions also seem to show support for a NBD model formulation.

〈Figure 1〉 An Empirical Distribution of Visit Frequency



The parameters (a and b) were estimated by solving the Mean and Zero equations simultaneously (Ehrenberg 1959; Morrison and Schmittlein 1988). The result of parameter estimation is: a = 0.049 and b = 0.109.

The results are summarized in <Table 1>.

From the Table, we can see how well the NBD estimated probability distribution fits the actual (or observed) probability distribution. The Chi-square value also shows that the two distributions are not significantly different. The null hypothesis that the underlying distribution of empirical frequency is a NBD distribution can be rejected because the Chi-squared value (4.87) is less than the critical value (8.672) with d.f. 17 and at the level of p-value=0.05. Thus, the Negative Binomial

〈Table 1〉 Evaluation of Fit of Visit Frequency Distributions

k	Actual	NBD	CNBD	$P_{actual}(k)$	$P_{nbd}(k)$	$P_{cnbd}(k)$
0	1056	1051.2	1053.9	0.897196	0.893145	0.895374
1	45	46.2	46.4	0.038534	0.039241	0.039423
2	21	21.8	17.2	0.016992	0.018553	0.014599
3	13	13.4	13.2	0.011045	0.011424	0.011235
4	10	9.2	9.5	0.008496	0.007850	0.008106
5	7	6.7	7.0	0.005098	0.005732	0.005988
6	7	6.3	6.1	0.007647	0.005348	0.005193
7	5	5.1	5.3	0.001699	0.004339	0.004508
8	4	3.8	4.7	0.003398	0.003191	0.003951
9	2	2.6	3.0	0.001699	0.002170	0.002532
10	2	2.1	1.9	0.005098	0.001771	0.001652
11	1	1.7	1.4	0.000000	0.001468	0.001202
12	1	1.4	1.3	0.000850	0.001211	0.001140
13	1	1.2	1.2	0.000000	0.001012	0.001055
14	0	1.0	1.1	0.000000	0.000850	0.000907
15	0	0.8	0.9	0.000000	0.000718	0.000800
16	0	0.7	0.8	0.000000	0.000609	0.000694
17	1	0.6	0.8	0.000850	0.000518	0.000698
18	0	0.5	0.6	0.000850	0.000443	0.000544
>=19	1	0.4	0.2	0.002549	0.000379	0.000369
Sum=	1177	1177.0	1176.7	1.000000	0.999861	0.999880
Chi-squared value		4.87	8.39			
d.f.		17	17			
P-value=Pr(X>=x)		0.9981	0.9574			
Estimated Mean (m)		0.402719	0.402719			
Parameter:						
a		0.048730	0.043284			
b		0.120809	0.214969			

distribution (NBD) model is a good approximation for consumer visit-frequency at a particular a Web site.

The estimated conditional probabilities are also summarized as follows: (1) New Triers= Pr(k≥1 in T2 given k=0 in T1) = 4.4% (2) Repeat Visitors = Pr(k≥1 in T1 and k≥1 in T2) = 8.7% (3) Repeat Visit Ratio = Pr(k≥1 T2 given k≥1 in T1) = 84.2%.

Managers can use these three measures to evaluate either site performances across different sites or marketing efforts such as advertising and promotional activities. If other things are assumed to be constant, it holds that the higher the penetration rate (% of new triers), the better the site. Using the same logic, we can say that the higher repeat visit ratio and the larger proportion of repeat visitors to total visitors, the better the site. Thus, managers can evaluate their promotional activities on the Internet by using these three measures.

2) CNBD Model Estimation

By using the Mean-and-Zero estimation procedure suggested in Chatfield and Goodhardt (1973), we got the following result: a = 0.043 and b = 0.215.

The results are summarized in Table 1, in there one can see that a CNBD model also provides a good fit for the observed frequency distribution. The Chi-squared value shows that the CNBD model distribution is not significantly different from the actual frequency distribution because Chi-squared value (=8.39) is less than the critical value (8.672) with d.f.=17 at the level of p-value=0.05. Thus, we cannot reject the null hypothesis that the actual frequency distribution follows a CNBD. In addition, this study confirms the finding in Chatfield and Goodhardt (1973) that both the CNBD model and the NBD model provide similar results and fit the empirical data very well.

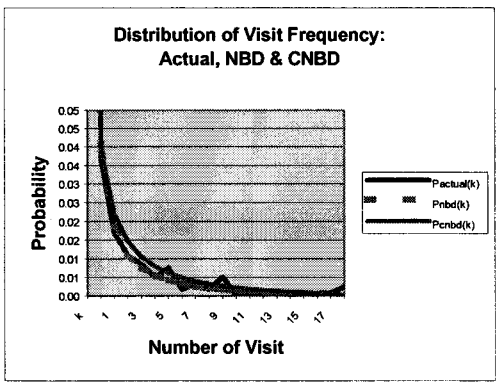
3) Evaluation of Fit of Visit Frequency: NBD vs. CNBD

The empirical investigation shows that the NBD model and the CNBD model produce similar estimates. The results are also summarized in Table 1. Although both models seem to provide reasonably good estimates of the actual frequency distribution, the NBD model is a slightly better compared to the actual consumer visit-frequency data (please refer to <Figure 2>), at Yahoo! in terms of Chi-squared values [Chi-squared NBD=4.87 < Chi-squared CNBD=8.39 <Chi-squared critical(17, 0.05)=8.672].

4) NB Regression Model Estimation

For the model estimation, we specify the following functional form of the Negative Binomial (NB) regression model for the empirical illustration. The functional form is:

〈Figure 2〉 Distributions of Visit Frequency: Actual NBD and CNBD



$$Log\lambda_i = \beta_0 + \beta_1 LAV\_TIME_{i1} + \beta_2 AN\_PAGE_{i2} \ldots + \beta_k GENDER_{ik} + \alpha_i$$

Where: λi = the expected value (mean) of the number of visits by consumer i

LAV_TIME = A logarithmic value of average time spent per visit on the Internet during a given period of time

AN_PAGE = Average number of pages visited per visit

GENDER = An explanatory variable (Male vs. Female)

We assume that (1) the dependent variable (yi) has a Poisson distribution with expected value (λi), conditional on ei, and that exp(ei) has a standard Gamma distribution (Allison 1999).

By using a SAS macro developed by Hilbe (1994), this study estimated the Negative Binomial regression. The results are given in <Table 2>. The results indicate that there are less visits to Yahoo! when a consumer spends more time browsing on the Internet in each session, while there are more visits to Yahoo! when the consumer has seen more pages per session. But the results show that there is no significant difference in the number of visits to Yahoo! between men and women. More specifically, for the variable (AV_PAGE), controlling for the other covariates, if consumers increase their average number of pages per session by one unit, then the number of visits to Yahoo! increases by 9.28% [100*(e0.0887 -

1)]. For LAV_TIME, controlling for the other covariates, the number of visits to Yahoo! decrease by 43.7% [100*(elog(0.2678) - 1)] as the consumer increases his/her average browsing time per session by one unit (the unit is 'second' for this study).

〈Table 2〉 Negative Binomial Regression for the Number of Visit

Log Negative Binomial Regression

Number of iterations:	11		
Alpha:	8.0691		
Deviance:	453.1981	Deviance/DF:	0.4146
Pearson Chi2:	1093.1076	Pearson Chi2/DF:	1.0001
LogLikelihood:	-942.9172		

OBS	PARM	DF	ESTIMATE	STDERR	CHISQ	PVAL
1	INTERCEPT	1	-3.3211	0.3698	80.6362	0.0001
2	LAV_TIME	1	0.2678	0.0514	27.1981	0.0001
3	AV_PAGE	1	0.0887	0.0242	13.3796	0.0003
4	GENDER	1	0.2147	0.2023	1.1266	0.2885
5	SCALE	0	1.0000	0.0000	.	.

Ⅳ. Managerial Implications

This study suggests the following managerial implications. To the extent that 'law-like regularities' govern consumers' repeat visits to a Web site, these regularities provide a standard by which to identify exceptional situations. Further, the prediction by the NBD model tells managers the extent of steady-state repeat visits without any explicit consideration on the effect of any promotional activities by the firms. Thus, if we compare the probability distribution including marketing activities such as advertising with this basic probability distribution without marketing activity, we can measure the pure effect of marketing activities on the consumers' repeat visit behavior. This can be investigated using a NB regression model with independent variables representing the marketing mix. The NB regression analysis will be discussed in a future study as an extension of this study.

In addition, the understanding of repeat visit behavior on the Internet has important implications for determining where to place advertisements. The manager as a media planner can improve the advertising effectiveness of the company's ads by choosing the best Web sites appropriate for the firm's advertising strategy. When the manager evaluates alternative sites for the company's advertisement, s/he can use the measures proposed in this study (i.e., the proportion of new triers, the proportion of repeat visitors, and the repeat visit ratio).

V. Conclusion

1. Summary of Findings

The study findings suggest two sets of

implications. The first involves a methodological issue. To date, the NBD model has been largely restricted to the behavioral context of brand purchasing, repeat buying or filler trips. The findings of this study suggest that the NBD model can be applied to a wider range of behavioral processes, including Web browsing.

The second implication refers to managerial issues. The prediction by the NBD model allows managers to measure the pure effects of marketing activities on consumers' repeat visit behavior. The independent variables representing the marketing mix are used within a NBD regression model. The effectiveness of the marketing variables can be evaluated. In addition, the understanding of repeat visit behavior on the Internet has important implications for determining where to place advertisements. The manager as a media planner can improve the advertising effectiveness of the company's ads by choosing the best Web sites appropriate for the firm's advertising strategy. When the manager evaluates alternative sites for the company's advertisement, s/he can use the measures proposed in this study (i.e., the proportion of new triers, the proportion of repeat visitors, and the repeat visit ratio).

The empirical results show support for the goodness of fit of the NBD model (See <Table 1>). In general, the probability distribution of consumers visiting a certain Web site displays the characteristics of the Negative Binomial distribution, as was the case in the studies of 'consumer purchasing behavior for the frequently purchased goods 'and' filler trip behavior. The probability distribution of k repeat visits to the Internet site tends to be uni-modal with a strong positive skew. The Chi-square test shows a reasonable correspondence between the actual ($P_{observed}$) and theoretical distributions (P_{NBD}) (Refer to <Table 1>).

2. Limitations and Future Research Opportunities

Despite the interesting findings of this study, there are many limitations mainly due to the data collection period (this study is based on the online panel data observed during only 4 weeks). One important limitation of this study is in the stationary property of the NBD model. That is, the NBD framework assumes a steady state condition. But in the real world, market situations are typically transient and thus vary over time. Therefore, the prediction by the model may not separate the effects of differences in the situations between T1 and T2.

The second limitation is due to the absence of a test of the model's predictive

accuracy. Such a test should be a major factor in evaluating the NBD model. Thus, future research should test the two models' predictive ability and compare them to the other alternative models.

Finally, this research did not investigate the effect of marketing efforts on visit frequency mainly due to data limitation. However, the model specification of the NB regression provides a framework that permits the incorporation of marketing variables. As more comprehensive online data, including marketing variables, becomes available, additional empirical studies with marketing mix variables as covariates, are needed to verify and to generalize the findings of this study in future studies.

Nevertheless, this study provides a natural extension of a NBD type of model to explain the consumer repeat visit frequency on the Internet. This study also suggests that continued research towards understanding and modeling consumer behavior on the Internet, as a function of explanatory variables (i.e., marketing mix variables, individual characteristics, site characteristics, and any other situational variables), should prove to be a fruitful endeavor.

REFERENCES

Allison, Paul D. (1999), *Logistic Regression Using The SAS System: Theory and Application*, SAS Institute Inc., Gary, NC, USA.

Chatfield, C. and G. J. Goodhardt (1973), "A Consumer Purchasing Model with Erlang Inter-Purchase Times," *Journal of the American Statistical Association*, Vol. 68, No. 344 (December), pp. 828-835.

DeGroot, Morris H. (1984), *Probability and Statistics*, 2nd ed., Addison Wesley, MA.

Dunn, R., S. Reader, and N. Wrigley (1983), "An Investigation of the Assumptions of the NBD Model as Applied to Purchasing at Individual Stores," *Applied Statistics*, 32 (3), 249-259.

Ehrenberg, A. S. C. (1959), "The Pattern of Consumer Purchases," *Applied Statistics*, 8 (March), 26-41.

Ehrenberg, A. S. C (1972), *Repeat-Buying: Theory and Applications*, New York: American Elsevier Publishing Company, Inc.

Engle, Robert (2000), "The Econometrics of Ultra-High Frequency Data," *Econometrica*, Vol. 68, No. 1, pp. 1-22.

Frisbie, Jr., Gil A. (1980), "Ehrenberg's Negative Binomial Model Applied to Grocery Store Trips," *Journal of*

Marketing Research, Vol. 17 (August), pp. 385-390.

Gupta, Sunil (1991), "Stochastic Models of Interpurchase Time With Time-Dependent Covariates," *Journal of Marketing Research*, 28 (February), 1-15.

Herniter, Jerome (1971), "A Probabilistic Market Model of Purchase Timing and Brand Selection," *Management Science*, Vol. 18, No. 4, Part II, December, pp. 102-113.

Hilbe, Joseph (1994), "Log Negative Binomial Regression as a Generalized Linear Model," *Technical Report 26,* Graduate College Committee on Statistics, Arizona State University, Tempe, AZ 85287.

Jeuland, Abel P., Frank Bass, and Gordon Wright (1980), "A Multibrand Stochastic Model Compounding Heterogeneous Erlang Timing and Multinomial Choice Processes," *Operations Research*, Vol. 28, No. 2, March/April, pp. 255-277.

Jones, Morgan and Fred Zufryden (1980), "Adding Explanatory Variables to a Consumer Purchase Behavior Model: An Exploratory Study," *Journal of Marketing Research*, Vol. 17, August, pp. 323-334.

Kim, Byung-Do and Kyungdo Park (1997), "Studying Patterns of Consumer's Grocery Shopping Trip," *Journal of Retailing,* Vol. 73(4), pp. 501-517.

Morrison, Donald G. and David C. Schmittlein (1981), "Predicting Future Random Events Based on Past Performance," *Management Science*, Vol. 27, No 9. September, pp. 1006-1023.

Morrison, Donald G. and David C. Schmittlein (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It worth the Effort?" *Journal of Business & Economics Statistics*, Vol. 6, No 2., April, pp. 145-166.

Schmittlein, David C., Albert C. Bemmaor, and Donald G. Morrison (1985), "Why does the NBD model work? Robustness in representing product purchases, Brand purchases and Imperfectly recorded purchases," *Marketing Science*, Vol. 4, No. 3, Summer, pp. 255-266.

Schumittlein, David C. and Donald G. Morrison (1983), "Predicting of Future Random Events with the Condensed Negative Binomial Distribution," *Journal of the American Statistical Association*, Vol. 78, pp. 449-456.

Schumittlein, David. C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science*, Vol. 33, No. 1, January, pp. 1-24.

Wheat, Rita D. and Donald G. Morrison (1990), "Estimating Purchase Regularity with Two Interpurchase Times," *Journal*

of Marketing Research, Vol. 27 (February), pp. 87-93.

Wu, Couchen and Hsiu-Li Chen (1999), "Counting Your Customers: Compounding Customer's In-Store Decisions, Interpurchase Time and Repurchasing Behavior," *Working Paper*, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C.

Zufryden, F. S. (1977), " A Composite Heterogeneous Model of Brand Choice and Purchase Timing Behavior," *Management Science*, 24 (October), 121-136.

Zufryden, Fred. S. (1978), "An Empirical Evaluation of a Composite Heterogeneous Model of Brand Choice and Purchase Timing," *Management Science*, Vol. 24, pp. 761-773.

Zufryden, Fred. S. (1991), "The WNBD: A Stochastic Model Approach for Relating Explanatory Variables to Consumer Purchase Dynamics," *International Journal of Research in Marketing*, Vol. 8, pp. 251-258.

〈한글초록〉

인터넷 이용자들의 웹사이트 재방문 빈도에 관한 실증적 연구

이 석 규*

본 연구는 소비자들의 선택모형에서 널리 사용된 NBD (Negative Binomial Distribution)타입의 계량적 모델 접근법이 온라인 상에서 소비자들이 특정한 기업의 웹사이트를 방문하는 행위를 설명하는데 적용될 수 있는지를 탐구한다. 본 연구에서는 다음의 두 가지 연구 주제를 다루고 있다. 첫째, 소비자들이 웹사이트를 반복하여 방문하는 행위의 빈도에 관한 분포를 확률적으로 규정하며, 둘째로는 그러한 소비자들의 반복된 이용빈도의 분포에 소비자들의 일반적인 인터넷 사용패턴과 인구 통계적인 변수들이 어떤 영향을 미치는 지를 조사하고 있다.

일련의 실증적 분석을 통하여, 이 논문은 마케팅의 선택모형 (Choice Model)들에서 널리 사용된 NBD 타입의 모델들이 인터넷상의 사이트 방문빈도 연구에도 잘 적용될 수 있음을 보여주고 있다. 그리고 이 연구는 이러한 소비자들의 이용빈도에 관한 모델개발이 온라인 기업의 당면문제에 어떠한 영향을 미치는 지를 설명한다. 특히 본 연구는 반복된 이용빈도와 소비자들의 일반적인 인터넷사용 특징 및 인구 통계적인 변수들과의 상호관계를 규명했다. 본 연구에서 제시된 모델들을 추정하고 검정하기 위해, 800,000번의 방문기록과 1000개 이상의 상이한 방문사이트 수로 구성된 웹 패널 데이터를 사용하여 실증분석을 연구에서 제시하는 모델을 개발하고 검증하였다.

주제어 : 웹사이트 방문행위, NBD 모델, 웹 패널 데이터분석, 사이트 충성도

* 한국외국어대학교 경영학과 조교수