Characteristics of Global Calling in VoIP services: A logistic regression analysis

Thanyalak Mueangkhot[1] , Julian Ming-Sung Cheng [2] and Chaknarin Kongcharoen [3]

[1] Department of Business Administration, National Central University, Jhongli City, Taoyuan Country, 32001, Taiwan (R.O.C.)

[2] Department of Business Administration, National Central University, Jhongli City, Taoyuan Country, 32001, Taiwan (R.O.C.)

[3] Department of Computer Science & Information Engineering, National Central University, Jhongli City, Taoyuan Country, 32001, Taiwan (R.O.C.)

Abstract

We investigate the characteristics of Voice over Internet Protocol (VoIP) services in Taiwan that influence Global Calling choice, by comparing Bayesian and Maximum likelihood estimates. The results will assist operators in setting their services. The sample data are 573 sales transactions using prepaid cards issued by Taiwanese e-retailers. The dataset contains individuals who placed recommendations in 2011-2013, which are representative of the actual market. Global Calling is defined as a critical characteristic selected by consumers. Data are analyzed by logistic regression of the two approaches. The study shows that the odds ratio of Global Calling is suitably modeled by the Bayesian approach. Collectively, the results show that the key characteristic of VoIP service is the price per month and Taiwanese landline calling. Taiwanese landline calling exerts the greater positive effect on Global Calling choice. The study then discusses the implications of VoIP operator selection, focusing on Taiwan's local call market.

Keywords: *Characteristic of VoIP, Logistic regression, Bayesian estimation, Maximum Likelihood estimation, Global Calling.*

1. Introduction

International calling (hereafter called Global Calling) is offered by Internet Network Operators, including the dominant international player Skype and others such as Phone Power, ITP, Lingo, Tokbox and iCall. Consumers can make international calls at zero or very low rates by Voice over Internet Protocal (VoIP). By the end of 2011, more than 2.3 billion people worldwide could access the internet [1]. To satisfy consumers, operators attempt to improve their own infrastructure technology and provide a range of services. The majority of VoIP call services utilize existing long distance calling, implementing voice technologies such as Global Calling.

Therefore, Global Calling is a critical characteristic by which operators can develop and increase revenue. Operators are interested in the service characteristics (attributes) that affect Global Calling services, and in which methods are applicable to small data sources. Using VoIP services, operators may strategically design and use instruments that promote Internet communication services.

In this paper, we investigate the impact of VoIP service characteristics on consumer decision in Taiwan's local call market. The accuracy of Bayesian predictors is compared that of Maximum likelihood (ML) estimates. The methodology uses the odds ratio of Global Calling. The estimated parameters provide insight into the characteristics of VoIP, and can thereby be used to improve international call performance.

The remainder of this study is organized as follows. The next section briefly reviews existing knowledge of VoIP services and the application of logistic regression to the service sector. Section 3 describes the data sources and evaluates the relationship between VoIP service characteristics and Global Calling choice by logistic regression. The results of the analysis are presented in Section 4, while Section 5 provides a summary and conclusions.

2. Literature review

The relevant literature can be broadly divided into management of the VoIP characteristics and application of the logistic model to the service sector. We discuss each class in turn below.

2.1 VoIP characteristics

Largely as a result of globalization, electronic communication technologies have become faster, cheaper, more reliable and more effective in recent decades, and have supplanted traditional means of sending information across the globe, which are slow and often unpredictable. VoIP has become an effective means of international calling at comparatively cheap rates. This innovation in the communication industry has led to new voice technologies, such as prepaid card calling. Prepaid phone cards are compatible with any phone, anywhere at any time, and have been adapted to VoIP services, where they are used to make international calls at reduced rates. These cards have led to rapid sales increases in commercial local markets such as voipindia.co.in, aglow.sg, indosat.com, skype.com, and skype.pchome.com.tw.

This study concentrates on the role of characteristic services in telecommunication sectors, which are perceived in different ways by consumers. Previously, the relationships between VoIP choice and telephone service features have been investigated by part-worth function models, which were validated on consumer data [2, 3].

In this study, VoIP characteristic services are defined from the packages or plans of VoIP services, which are presented in promotional messages on operator websites. Examples are:

Skype.com: Unlimited calls, Landline calls, Mobile calls, Landline and mobile calls, credit payment, subscription payment.

aglow.sg: international calls, traditional analog phone lines, mobile phone line.

Skype.pchome: Unlimited world, Taiwan landline calls, Taiwan mobile calls, Unlimited United States calls, China calls.

Most of the VoIP operators offer similar service plans with various features. This research uses Skype as a representative VoIP service, on account of its current popularity in voice networking. Skype enables international calling through any landline or mobile phone, including received calls and voicemail messages. The main function of Skype is to support long distance and local calling by consumers.

2.2 Logistic model

Many techniques have been proposed for analyzing the characteristics or attributes of telecommunication products [2,3,4]. However, comparatively few studies in the VoIP service literature have related service characteristics to VoIP consumer choice. Here, we evaluate a VoIP transaction dataset by Bayesian logistic regression, and compare the results with those estimated by ML. ML is a preferred analysis technique because its estimators are consistent, sufficient, and efficient [5]. The parameters estimated by ML and Bayesian regression are compared in terms of their predictive accuracy in telecommunication sectors.

3. Research Methodology

In our methodology, the variables are VoIP service characteristics. We model the relationship between these service characteristics and Global Calling service choice.

3.1 Data

This data are e-retailer sales transaction data on prepaid cards sold in Taiwan. The dataset includes information on retail choice by consumers who had purchased Skype prepaid cards and had recommended used services in blogs. Many Taiwanese customers purchased from two e-retailers on the Yahoo (yahoo.com.tw) and Ruten (ruten.com) sites. The dataset comprised 573 individual records logged in 28 months between 2011 and 2013 of the service plan. We identified 9 service characteristics among 8 VoIP service plans from messages promoting prepaid cards on e-retailers’ websites. These are reported in Table 1.

In summary, the data consists of 9 service characteristics as independent variables for 8 VoIP packages. The data are categorized as Global Calling or non-Global Calling as dependent variables. These are defined as the voice characteristics which influence consumer choice. Data collection is implemented from the service features advertised in the prepaid card messages. We selected 8 service plans which could be purchased at the time of data collection.

3.2 Logistic regression approach

Logistic regression is used to predict a discrete outcome. The dependent variable is dichotomous, such as presence/absence or success/failure. When the independent variables are categorical or mixed continuous/categorical, or when the dependent variable is dichotomous, logistic regression is the preferred analysis technique.

Table 1 Variables and descriptions of VoIP service characteristics used in the analysis

Variables	Definition	Mean	S.D.
GLOBAL	Plan call listing included call to landline and mobile for 10 countries (Canada, China, Guam, Hong Kong SAR, Puerto Rico, Singapore, Thailand, United States) and call to landline for 65 countries (available at www.Skyp.com/en/rates/)	0.430	0.495
TW_JP_KR	Plan call listing included call to Taiwan, Japan and Korea	0.070	0.252
US_CA	Plan call listing included call to USA and Canada	0.010	0.072
TW_L_UN	Plan call through Taiwan landline unlimited included	0.790	0.407
TW_L	Plan call through Taiwan landline included	0.990	0.110
TW_M	Plan call through Taiwan mobile included	0.482	0.501
ONE_M	price prepaid per one month ($TWD)	273.770	95.303
THREE_M	price prepaid per three months ($TWD)	733.170	279.707
ONE_YEAR	price prepaid per one year ($TWD)	2625.630	837.394
OLD_CUS	Plan for old consumers	0.090	0.290

1.000 Taiwan Dollar ($TWD) = 0.033 $ US Dollar

In logistic regression, the relationship between the predictor variable and dependent variables is non-linear. The logistic regression function is the logarithmic transformation of the probability, and is defined as

$$\theta = \frac{e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}}{1+ e^{(\alpha+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}} \quad (1)$$

where x_i are the predictors, α is a constant and, β is the coefficient of the predictor variables.

This transformation yields a number, called logit P(X), for each test case with input variable X. The logistic regression model linearly fits the log odds to the variables [6].
An alternative from of the logistic regression equation is given as

$$log\left[\frac{\theta}{1-\theta}\right] = logit\ P(X)$$

$$= \alpha \ + \textstyle\sum_{i=1}^{k} \beta_i X_i$$

$$= \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (2)$$

This study compares the two approaches of ML estimation and Bayesian estimation.

The first approach is a classical statistical method of parameter estimation. ML estimates the unknown parameters in a logistic regression model [7]. ML maximizes the log likelihood, defined as the likelihood (odds) that the observed value of the dependent variable will be predicted from the observed values of the independents. The results of logistic regression are expressed in terms of odds rather than probabilities because the odds provide a pure summary statistic of the partial effect of a given predictor, controlling for the predictors in the logistic regression.

The second approach is a Bayesian statistical method of parameter estimation. The Bayesian regression model uses the Markov Chain Monte Carlo (MCMC) approach which is simple and accurate method [8]. MCMC involves three steps: (a) formulating prior probability distributions of targeted parameters; (b) specifying the likelihood function; and (c) MCMC sampling of the posterior probability distributions. In the Bayesian approach, parameters are considered random and a joint probability model for both data and parameters is required. This difficulty is most easily circumvented by proposing an informative prior of small precision, to avoid potential bias introduced by subjective beliefs [9]. Because the joint posterior distribution of the parameters of the proposed models is analytically intractable, the point and interval estimates of the parameters are obtained by simulation-based MCMC methods [10].

After estimating a suitable prior distribution of the model parameters, the model runs simultaneously in WinBUGS. The parameters are defined from the data by a set of vague normal priors. A burn-in of 5,000 iterations is permitted, followed by 10,000 iterations in which the calculated intercept and coefficients are stored. Convergence was achieved after 5,000 iterations and the posterior distributions of model parameters were summarized using descriptive statistics. ML estimator of the model parameters was obtained in R, using the Rcmdr package introduced and developed by John Fox [11]. Comparison of the obtained results was also implemented in this package.

In this study, the dependent variable is Global Calling—*Global*—, which takes value 1 if the consumer has used the Global Calling service and 0 otherwise. The independent variables are *TW_JP_KR*, *US_CA*, *TW_L_UN*, *TW_L*, *TW_M*, and *OLD_CUS* is a dummy indicator variable. The three cost-per-period levels, *ONE_M*, *THREE_M*, and *ONE_YEAR*, are continuous variables. The analysis attempts to highlight the relationship between the VoIP service characteristics and Global Calling choice. A logistic regression model can be written in terms of the odds of an event occurring. These odds are defined as the ratio of the probability that an event will occur and the probability that it will not. The coefficients indicate the effect of individual explanatory variables on the logarithm of the odds. In terms of the variables used in this study, the logistic regression equation is

$$\mathrm{Y_i} = \alpha \ +\beta_1(TW\_JR\_KR) + \beta_2(US\_CA) + \beta_3(TW\_L\_UN) + \beta_4(TW\_L) + \beta_5(TW\_M) + \beta_6(ONE\_M) + \beta_7(THREE\_M) + \beta_7(ONE\_YEAR) + \beta_8(OLD\_CUS) + \varepsilon_i \quad (3)$$

where *Yi* is the log odds of Global Calling for the *i*th observation and ε_i is its error term. As shown in Eq. (3), the probability of Global Calling is affected by the difference between Global Calling and its absence. In this study, the parameters are estimated by both Bayesian and ML methods.

4. Results

The results are summarized in Tables 2 and 3. The fit models of the two approaches are assessed and the estimates are reported. Our main interest lies in analyzing the relationship between the presence/absence of Global Calling in VoIP service plans and the various characteristics of the service.

4.1 Model fitting

The model is based on Bayesian and ML estimation. The eight predictors are the characteristics of the VoIP service extracted from the promotional messages of prepaid cards posted on company websites. We compared the model with respect ot Akaike’s Information Criterion [12] and Deviance Information Criterion [13]. The DIC value reported by Bayesian analysis was 204.8, while the AIC reported by ML was 213.89. The smaller DIC value indicates that the Bayesian method provides a better fit to the data.

4.2 Variable effects

The estimates output by the two approaches (Bayesian logistic regression and ML) are summarized in Tables 2 and 3. The results involve eight predictors and the outcome is the probability of *GLOBAL*.

In the Bayesian method, the posterior estimations of beta coefficients are obtained as the means of simulated betas from different numbers of samples after confirming convergence. On the other hand, the ML estimates of beta coefficients are obtained by iterating the Newton–Raphson method on the data. In both cases, the results are most conveniently explained in terms of the odds ratio (the exponential of the betas coefficients). The respective posterior credible intervals (or confidence intervals) can be readily obtained in the usual manner.

The variables considered are dummies, discrete or continuous variables. The odd ratio of a discrete variable can be interpreted as the change in relative risk of a certain outcome associated with a change in the independent variable. Values below 1 imply a negative association and their magnitude indicates their contribution to the probability of the outcome. The odds ratios for the continuous variables, on the other hand, can be interpreted as elasticities; that is, the odds ratio indicates the extent to which a one per cent change in an independent variable affects the relative risk of the outcome. As for the discrete variables, values below 1 imply negative elasticities. This study presents the odd ratio analysis of the discrete variables in *GLOBAL* and non *GLOBAL.*

The Bayesian analysis indicates that five variables are significantly correlated with Global Calling. Two variables, *ONE_M* and *TW_L*, make significantly positive contributions. On the other hand, ML analysis reveals only two variables that contribute significantly to Global Calling choice, of which one (*ONE_M*) makes a positive contribution. *THREE_M* exerts a significantly negative effect on the service.

The odds ratio, together with the respective 95% posterior credible intervals and 95% confidence intervals, are also presented in Tables 2 and 3. Similar results are obtained for both Bayesian and ML estimation methods. The *ONE_M* variable is the characteristic yielding the highest number of sales. *THREE_M* and *ONE_YEAR* are variables of higher cost than *ONE_M* and generate fewer sales than *ONE_M*. From these results, it is evident that consumers who select higher-cost services are less likely to intensify the service. The results are supported by both Bayesian and ML at the 5% significance level.

The *TW_L* variable exerts a positive impact on Global Calling. In the real market, if an operator enables the *TW_L* characteristic, the volume of sales is likely to

Table 2: Bayesian Estimates in Logistic model

Parameter	mean	2.50%	97.50%	Odd Ratio	
alpha	-3.43E + 01	-44.850	-21.120	1.33E - 15	
ONE_M	1.10E - 01	0.073	0.146	1.12E + 00	*
THREE_M	-2.17E - 02	-0.032	-0.014	9.79E - 01	*
ONE_YEAR	1.02E - 03	-0.002	0.004	1.00E + 00	
TW_L	2.42E + 01	8.336	35.900	3.20E + 11	*
TW_M	-1.00E + 02	-132.100	-70.770	2.76E - 44	*
TW_L_UN	-6.21E + 00	-15.200	-0.013	2.01E - 03	*
US_CA	-1.31E + 01	-64.830	22.900	2.02E - 06	
TW_JP_KR	-4.93E + 00	-63.740	44.250	7.21E - 03	
OLD_CUS	8.06E - 01	-1.216	2.892	2.24E + 00	

Signif. Code: '*'

Table 3: Maximum Likelihood Estimates in Logistic model

variable	Estimate	Std. Error	Pr(>|z|)	Odd Ratio	
Intercept	-4.41E + 01	7.44E + 03	0.995	7.33E - 20	
ONE_M	1.49E - 01	2.68E - 02	3.10E - 08	1.16E + 00	*
THREE_M	-2.41E - 02	5.46E - 03	9.92E - 06	9.76E - 01	*
ONE_YEAR	-1.54E - 03	1.81E -03	0.393	9.98E - 01	
TW_L	4.52E + 01	1.24E + 04	0.997	4.22E + 19	
TW_M	-7.54E + 01	9.93E + 03	0.994	1.78E - 33	
TW_L_UN	-1.94E + 01	9.88E + 03	0.998	3.61E - 09	
US_CA	1.16E + 01	1.27E + 04	0.999	1.09E + 05	
TW_JP_KR	-2.71E + 01	2.91E + 04	0.999	1.65E - 12	
OLD_CUS	1.03E + 00	1.59E + 00	0.517	2.79E + 00	

Signif. Code: '*'

increase. Therefore, the operator is likely to design service plans that support international calling. These conclusions can be drawn only from the Bayesian analysis; ML estimation did not identify a similar impact of the *TW_L* variable.

A primary goal of the present study is to compare the cause-effect estimates from ML and Bayesian analysis. Interestingly, when the two methods were applied to Global Calling data, they yielded different results. The variables *ONE_M*, *THREE_M*, *TW_L*, *TW_M* and *TW_L_UN* showed a cause-effect relationship under Bayesian estimation, while ML estimation implicated just two variables, *ONE_M* and *THREE_M*, in the cause-effect of VoIP services. The results also demonstrate the applicability of the Bayesian approach to sample observations.

5. Conclusions

The purpose of this study was to predict the important variables that influence Global Calling choice probability in the Taiwanese local call market. To this end, we identified the VoIP characteristics that most strongly affect Global and non-Global Calling by logistic regression analysis using Bayesian and ML methods. The methods were compared in terms of their AIC and DIC statistics. The results show that *ONE_M* and *TW_L* make significantly positive contributions to *GLOBAL* choice, while *TW_M*, *TW_L_UN* and *ONE_YEAR* exert a negative influence. This study has also identified the Taiwan landline service package as the greatest positively contributing factor.

The Bayesian method provided a better fit to the data than the ML method, and identified a larger number of significant factors. Similar statistical studies will assist operators in elucidating VoIP service patterns. Furthermore, researchers can identify interesting transaction data that will assist theoretical development in communication technology.

The empirical results of this study have implications for Internet network operators. The main factors predicted to influence customer choice in Global Calling services depend on the adopted analysis approach. We conclude that the Taiwanese local call market should retain the landline feature in its service package.

Acknowledgments

Special acknowledgment is extended to Charnarin for assistance with data collection and for the many valuable discussions. We also acknowledge Ryan Chen for motivating this study, and who instigated useful discussions on the need to identify the characteristics of VoIP services from commercial Taiwanese websites.

References

[1] ITU, "Internet continues to grow Worldwide to reach 2.3 billion mark late June 2012", available at: http://www.itu.int. , accessed 5 May 2013.
[2] F.m Tseng, "Forecasting the Taiwan customer market for Internet telephony", Journal of the Chinese Institute of Industrial Engineers, Vol. 22, No. 2, 2005, pp. 93-105.
[3] M. L. Zubey, W. Wagner, and J. R. Otto, "A conjoint analysis of voice over IP attributes", Internet Research, Vol. 12, No. 1, 2002, pp. 7-15.
[4] G. Cecere, and N. Corrocher, "The usage of VoIP services and other communication services: An empirical analysis of

Italian consumers", Technological Forecasting and Social Change, Vol. 79, No. 3, 2012, pp. 570-578.
[5] M.E.B. Akiva, and S. R. Lerman, "Discrete choice analysis: theory and application to predict travel demand", MIT press, Vol. 9, 1998.
[6] B. W. Lindgren, Statistic theory, Macmillan College, 1976.
[7] S.E. Fienberg, The analysis of cross-classified categorical data: Springer, 2007.
[8] R. S. Akhter, S. Hussain and U. R. S. Habib, "MCMC simulation of GARCH model to forecast network traffic boad", International Journal of Computer Science Issues, Vol. 9, No.2, 2012, pp. 277-284.
[9] R. Haylock, and A. O'Hagan, "On inference for outputs of computationally expensive algorithms with uncertainty on the inputs", Bayesian statistics, Vol. 5, 1996, pp. 629-637.
[10] L. Tierney, "Markov chains for exploring posterior distributions", The Annuals of Statistics, 1994, pp. 1701-1728.
[11] J. Fox, "Getting Started With the R Commander: A Basic-Statistics Graphical User Interface to R", Journal of Statistical Software, Vol. 14, No. 9, 2005, pp. 1-42.

[12] H. Akaike, "Information theory and an extension of the maximum likelihood principle", Selected Paper of Hirotugu Akike, 1998, pp. 199-213.
[13] D.J. Spiegelhalter, N. G. Best, B. P. Carlin et al., "Bayesian measures of model complexity and fit", Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 64, No. 4, 2002, pp. 583-639.

Thanyalak Mueangkhot received her M.B.A in Faculty of Liberal Arts and Management of Kasetsart University, SakonNakhon campus, Thailand, in 2005. She is now a graduate in Department of Business Administration of National Central University. Her main research interest is service marketing, green marketing, and Internet Technologies.

Julian Ming-Sung Cheng is an Associate Professor of Marketing in the Department of Business Administration, National Central University, Taiwan. His current research interests include marketing channels, international branding, global marketing and internet marketing. His papers have appeared in Journal of International Marketing, Industrial Marketing Management, Journal of Business and Industrial Marketing, Journal of Advertising Research, International Journal of Market Research, Journal of Product & Brand Management, Journal of Marketing Channels, Asia Pacific Journal of Marketing and Logistics and so on.

Chaknarin Kongcharoen received his M.Eng in Computer Engineering in Faculty of Engineering, Suranaree University of Technology, Thailand, in 2006. He is now a graduate in Department of Computer Science & Information Engineering, National Central University. His main research interest in data mining, web mining, software metric and e-learning. His papers have appeared in KMITL Science and Technology Journal.