Factors Influencing Tourism Demand to Revisit Pha Ngan Island Using Generalized Maximum Entropy

Apiwat Ayusuk[†,1] **and Kittawit Autchariyapanitkul**[‡]

[†]Department of Economics, Faculty of Liberal Arts and
Management Sciences, Prince of Songkla University
Surat Thani Campus, Surat Thani, Thailand
e-mail : apiwat.a@psu.ac.th

[‡]Faculty of Economics, Maejo University, Chiang Mai, Thailand
e-mail : kittawit_a@mju.ac.th

Abstract : This study aims to examine the predictive factors of a tourist's decision to revisit Pha Ngan Island in Surat Thani Province of Thailand. A sample survey was collected from a random sample of tourists to Pha Ngan Island. The binary logistic regression model was applied. We use the method of generalized maximum entropy for estimating the model parameters. Based on the accuracy criteria of mean square error(MSE) to justify the model, the present study demonstrated that the prediction of tourism demand for revisiting Pha Ngan Island which is obtained from logistic regression by using GME method is more accurate than ML method, in the case of a small sample size. We also found that price and promotion are the factors influencing to tourism demand.

Keywords : choice; entropy; logit; logistic; tourism demand; Pha Ngan.

2010 Mathematics Subject Classification : 62P20; 91B84.

---

[1]Corresponding author.

1   Introduction

[1] purposed the modern econometric model for estimating the tourism demand and mentioned the research trends related to tourism demand forecasting: "Many researchers involved with quantitative causal tourism modeling and forecasting to compute many regression equations and try to find the finest model". Most of the variables for the tourism demand in the choice model could be a qualitative variable. [2] has developed a logistic regression for a binary response to analyze the data such as an event could happened or could not happened, chosen or not chosen. [3] used a logit model to calculate the probability of Portuguese tourists for choosing Brazil as a tourism destination. The estimated value of the dependent variable of the model will be the probability of occurrence of a value ranging between $[0, 1]$.

Tourism is considered as a major economic sector that has contributed to the economic growth of Thailand. The tourism demand in Thailand has been growing steadily for the last decade. Tourism industry benefits for related businesses, for instance, travel agencies, hotels, restaurants. In 2015, tourism industry impacts 5.8% of GDP of Thailand. Most of the tourists are from China, Russia, Malaysia, UK and Australia according to data from the Department of Tourism, Ministry of Tourism and Sport. Many papers on tourism demand such as [4], [5], [6], [7]. In their studies, they applied econometric techniques to determine the elasticity of tourism demand and to forecast tourism demand. In this study, we attempt to illustrate the factor affecting revisiting decision on tourism demand in Pha Ngan Island by applying a logistic regression with a concept of maximum entropy.

The classical logistic model is a parametric statistic, which gives the mean and standard deviation of the population based on distribution assumption and uses the maximum likelihood (MLEs) principle as an estimation method. However, The MLEs are solutions of maximization problems. But maximization problems might not have solutions. In such cases, we have to look for other methods of estimation. The point is this, in nice situations, MLE is a popular way to find good estimators. But it is by no means that it is a universal method of estimation. An alternative method for estimation comes from the modern information theory in which a mathematically robust measure of probabilistic information, namely, the entropy, has been developed.

The main idea behind the ME principle is this. We should select a probability distribution which is consistent with our knowledge and introduce no unwarranted information. Any distribution (satisfying known constraints) which has smaller entropy will contain more information (less uncertainty) and hence says something stronger than what we know. The distribution with ME (satisfying our known constraints) is the one which should be least surprising in terms of the predictions it makes. The ME principle guides us to the best distribution which reflects our current knowledge and it tells us what to do if experimental data do not agree with predictions coming from our chosen distribution.

The ME is useful in a variety of situations (e.g., in Econometrics). For decision-making as in a flooding prevention project, the choice of joint distributions (via,

copulas, say) based on ME has several advantages:

(i) it incorporates as much (or as little) information as there is available, (ii) it makes no assumption on a particular form of the joint distribution, (iii) it can be applied to both numerical and qualitative random variables (as it only involves the distributions and not the "values" that the random "elements" take, for example, categorical variables such as those take "values" as "low,middle, high"), (iv) it can take into account of any form of constraints, not only moments and linear correlations.

Finally, note that the rationale for choosing the most likely distribution in the light of the observed data by ME is somewhat similar to the *Maximum Likelihood Principle (MLP)* that you are all familiar with. Recall that, in the context of MLP, the "rational" estimator (function of the data) of a population parameter is taken as the argument of the maximization problem $\max L(\theta|X_n)$ over $\theta \in \Phi$, i.e., the statistics $\theta(X_n)$ making the "likelihood" (= probability of observing the data X_n) maximal. Here, it is the "entropy" which replaces the likelihood. The maximum entropy distribution is the most likely distribution giving rise to an observed state since it can generate more states than any other distributions, and as such, the observed state is more likely to belong to its set of possible outcomes.

We are "estimating" the whole distribution function: it is a nonparametric problem. In the sequent, we will elabore more on the role of data in the entropy maximization problem.

This paper is structured as follows. In Section 2 exhibits the theoretical background of GME logit model, while section 3 explains the empirical application and a survey data of tourist. Section 4 reports the empirical results and discussion, and final section gives conclusions.

2 Theoretical Background

In this section we give a brief conception of the GME logit and the formulation of the model and how the estimation method is working.

2.1 Generalized Maximum Entropy Logit

Discrete choice model such as logit is wildly used in the research. This model the probability (p) that individual i will choose or deal with a precise outcome from the set of possible outcomes [8]. This estimation method require the probability that an outcome is observed, and they impose that the most likely outcome is the one observed [8]. However, logit assume a parametric structure on the probabilities. The underlying distribution for the probabilities is unknown, and the choice of logit depends on this strong assumption [9]. Thus [9] and [10] favor to use estimation approaches that do not commit to such a strong assumptions.

Maximum entropy involves detecting the probability distribution (p_i) for the set of random variables (x_i). Generalized maximum entropy (GME) based on maximum entropy by including intuitive innovation terms. For the GME result,

we examine a study including T trials. In each case with J unsequenced feasible outcomes, a binary set of values y_{tj} is observed, where y_{ij} for $i = 1, 2, ..., T$ takes one of the J unsequenced groups $j = 1, 2, ..., J$. On each trial i, each alternative J is observed in the format of a binary variable, y_{ij}, that equals one if alternative j is chosen and zero otherwise. Let p_{ij} be the probability of choice j on trial i and be associated to a set of variables through the equation

$$p_{ij} = \text{Prob}(y_{ij} = 1|x_i, \alpha_j) = F(x_i'\alpha_j) > 0 \quad \text{for all i and j} \tag{2.1}$$

where α_j is a $(K \times 1)$ vector of unknowns, x_i' is a $(1 \times K)$ vector of predictive variables, and $F(\cdot)$ is a function uniting the probabilities p_{ij} with the covariates $x_i'\alpha_j$ such that $\sum_j F(x_i'\alpha_j) = 1$. The likelihood function shows the values of α in terms of know, fixed values for y. Thus,

$$L(\alpha|y) = \prod_{i=1}^{N} \frac{n_i!}{y_i!(n_i - y_i)!} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \tag{2.2}$$

The maximum likelihood calculates are the values for α that maximize the likelihood function in (2.2). Then, We allow the model to have noisy data, then the equation is

$$y_{ij} = F(\cdot) + e_{ij} = p_{ij} + e_{ij} \tag{2.3}$$

where p_{ij} indicates the unknown multinomial probabilities and e_{ij} represents the natural noise elements for each observation and is accommodated in the $[-1, 1]$ support space for each observation.

To find out the unknown p and undetectable e, we use the known covariates, x_i and the noisy observable, y_{ij} to formulate the problem. This information is used as the cross-moments between the x matrix and all the other variables

$$(I_j \otimes X')y = (I_j \otimes X')p + (I_j \otimes X')e \tag{2.4}$$

where X is a $(T \times K)$ matrix and y_{ij} is a $T \cdot J$ observations. Thus, a vectorized y_{ij} are going to be $(T \cdot J \times 1)$ vectors. The same employs for e and p, which are also vectors of dimension $(T \cdot J \times 1)$. From equation (2.3), it is an ill-posed problem, a problem which may have more than one solution where there are $\{T \times (J-1)\}$ unknown parameters but only $(K \times J) < \{T \times (J-1)\}$ data points.

To apply Shannons entropy measure to formulate the problem, we need to reparameterize the noise terms. Due to p is already in a probability form, only the components of e need to be reparameterized to continue with GME, as introduced by [9]. Because each e_{ij} will range between $[-1, 1]$, we are including a set of discrete points (v_{ij}) ranging between $[0, 1]$. The error terms are characterized by an H-dimensional support space, v, and w as a weights of an H-dimensional vector that correspond to each v. The unknown weights have the properties of probabilities with $\sum_h w_{ijh} = 1$. Thus the resulting reparameterization is

$$e_{ij} \equiv \sum_h v_{ijh} w_{ijh}, \tag{2.5}$$

where the H-dimensional errors support recommended by [10] as a traditional choice is $v = (1/\sqrt{T}, 0, 1/\sqrt{T})$ for each e_{ij}. Thus, In this paper, $H = 3$. The equation (2.3) takes the form

$$(I_j \otimes X^{'})y = (I_j \otimes X^{'})p + (I_j \otimes X^{'})(wv), \qquad (2.6)$$

where w is a $T \cdot J \times H$ matrix, and v is an $H \times 1$ matrix.

As above, the solutions for ML and ME multinomial logit are equivalent, regardless of their formulations are distinction (see [8]). As [11] argue, this parallel can be described by the following: 1) the estimating equations or moment constraints in the ME formulation are the ML logit first-order conditions; and 2) the ME solution resulting from the optimization has the same form as the logistic multinomial probabilities. As mentioned in [12] the GME dominate the ME because the GME is a product of two logits for the p and for the w, (see [9]).

3 Methodology and Data

3.1 Model Formulation

The functional form of the GME discrete choice model is a dual objective function. The dual objective function is composed of the entropy of the probabilities (p) and the entropy for the weights (w) (see [9], [10]). This implies the assumption of independence between the two. The objective function of the GME multinomial problem is the maximization of the Shannon entropy measure and takes the following form specified in [9, 13, 14],

$$\max_{p,w} H(p, w) = \max_{p,w} (-p^{'} \ln p - w^{'} \ln w) \qquad (3.1)$$

subject to the JK information-moment conditions

$$(I_j \otimes X^{'})y = (I_j \otimes X^{'})p + (I_j \otimes X^{'})(wv) \qquad (3.2)$$

the normalization constraints,

$$(I_{T1}I_{T2}...I_{TJ})p = 1 \quad \text{for} \quad i = 1, 2, ..., T \qquad (3.3)$$

where h is a $(1 \times H)$ vector of 1s.

The corresponding Lagrangian is

$$L = -p^{'} \ln p - w' \ln w$$
$$+ \lambda^{'}\{(I_J \otimes X^{'})p + (I_J \otimes X^{'}V_w - (I_J \otimes X^{'}y\}$$
$$+ \mu^{'}\{h - (I_{T1}I_{T2}...I_{TJ})p\} + p^{'}(1 - h^{'}w)$$

where λ, μ, and ρ are the corresponding Lagrange multipliers. Under this specification, the parameters of interest are λ. Note that these provide the coefficients

$-\lambda_j = \beta_j$. To simplify the math, we proceed with the Lagrangian in scalar form,

$$\max_{p,w} H(p, w) = \max_{p,w} \left(-\sum_{ij} \ln p_{ij} - \sum_{ijl} w_{ijl} \ln w_{ijl} \right) \quad (3.4)$$

subject to the JK information-moment conditions [9, 13], where the JKth condition is

$$\sum_i y_{ij} x_{ik} = \sum_i x_{ik} p_{ij} + \sum_{il} x_{ik} v_l w_{ijl} \quad (3.5)$$

When $\sum_j p_{ij} = 1$ and $\sum_l w_{ijl} = 1$

Note that $\beta_{jk} = -\lambda_{jk}$ and the first-order conditions give us

$$p_{ij} = \exp\left(-1 - \mu_i - \sum_k \lambda_{jk} x_{ik} \right)$$

The summation of probability equation over j and the summation of the noise term weights over h are

$$\sum_j p_{ij} = \exp(-1 - \mu_i) \sum_j \exp\left(-\sum_k \lambda_{jk} x_{ik} \right) = 1$$

and

$$\sum_l w_{ijl} = \exp(-1 - \rho_i) \sum_l \exp\left(-\sum_k \lambda_{jk} x_{ik} v_l \right) = 1$$

Then, a zero $(K \times 1)$ vectors are set, given $\lambda_1 = 0$, we get

$$\hat{p}_{ij} = \frac{\exp\left(-\sum_k \hat{\lambda}_{jk} x_{ik} \right)}{1 + \sum_{j=2}^{J} \exp\left(-\sum_k \hat{\lambda}_{jk} x_{ik} \right)}$$

and

$$\hat{w}_{ijl} = \frac{\exp\left(-\sum_k \hat{\lambda}_{jk} x_{ik} v_l \right)}{\sum_h \exp\left(-\sum_k \hat{\lambda}_{jk} x_{ik} v_l \right)}$$

We can show log odds-ratios as same as the traditional logit, by calculate from the equation below [11]

$$\ln\left(\frac{p_{ij}}{p_{i1}}\right) = -x_i \lambda_j \quad (3.6)$$

Coincidentally, we illustrated the odds ratios as the exponentiated logistic regression coefficients. The dual unconstrained formulation of the GME problem denotes that the GME is associate with a class of generalized logit foundations (see, [9];

[10]). Starting from Lagrangian, the dual unconstrained GME formulation as a function of the Lagrangian multipliers, λ, is, according to [9] and [10]

$$M(\lambda_j) = y'(I_j \otimes X)\lambda + \sum_i \ln\{\Omega_i(\lambda_j)\} + \sum_i \sum_j \ln\{\Psi_i(\lambda_j)\} \qquad (3.7)$$

We can minimize the dual objective function subject to λ, we get the multipliers $\hat{\lambda}$. From these, we get the values of $\hat{P}_{ij}$ and $\hat{P}_{jh}$ as well. The error components converge to zero as the sample size increases.

3.2 The Evaluation Criterion

In this study, we used the mean square error (MSE) as a criterion to evaluate the accuracy of the prediction in each estimation method. The MSE can be written as

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_t)^2, \qquad (3.8)$$

where N is a number of observation. $(Y_i - \hat{Y}_t)^2$ represents the error from our predictions.

3.3 Data

Pha Ngan Island is one of the most famous tourist attractions in Suratthani province, southern of Thailand. The main activity for visitors here is the Full Moon party. It is very well known for the world class travelers, held every evening under the full moon. Hence, this study is empirically collected the primary data in 2017 by applying the survey method with questionnaires to 400 tourists (N = 400) in Pha Ngan Island. We used questionnaires for determining the values, demographics, behaviors, attitudes, beliefs, and other information relating with a group of tourist. Once we got the completed data.

4 Empirical Results and Discussion

The results will be evaluated using advanced econometric techniques in order that we describe the relationship between marketing mixes, full moon party and revisit intention of tourists. Advanced econometric techniques focus on logistic regression, which employ the maximum likelihood and generalized maximum entropy to be the procedure for finding the value of parameters. The logistic regression can be defined by a function as follows:

$$REV_i = \alpha_0 + \alpha_1 FM_i + \alpha_2 PD_i + \alpha_3 PR_i + \alpha_4 PL_i + \alpha_5 PM_i + \varepsilon_i \qquad (4.1)$$

where $REV_i = 1$ if tourists have intention to revisit, $REV_i = 0$ otherwise. α_i are the regressors. ε_i is the error term. FM_i is dummy variable, $FM_i = 1$ if full

moon party is the motivation to travel in Pha Ngan Island, $FM_i = 0$ otherwise. This study designs the satisfaction of the marketing mix variables of the tourists coming to Pha Ngan Island. There are product PD_i, price PR_i, place PL_i and promotion PM_i, these are measured in likert scale.

Table 1: Parameter estimates of logistic regression model for tourist demand

N	100		200		400	
Coeff.	ML	GME	ML	GME	ML	GME
α_0	-11.029	-7.870	-3.512	-3.212	-3.206	-3.075
	$(3.846)^{***}$	$(2.866)^{***}$	$(1.618)^{***}$	$(1.541)^{***}$	$(1.141)^{***}$	$(1.118)^{***}$
α_1	-0.028	0.265	-1.055	-0.999	-0.154	-0.151
	(1.245)	(1.093)	(0.543)	(0.529)	(0.309)	(0.304)
α_2	0.402	0.270	0.203	0.188	0.279	0.267
	(0.734)	(0.625)	(0.432)	(0.417)	(0.298)	(0.293)
α_3	1.729	1.315	0.800	0.744	0.636	0.617
	$(0.557)^{***}$	$(0.434)^{***}$	$(0.288)^{***}$	$(0.277)^{***}$	$(0.190)^{***}$	$(0.187)^{***}$
α_4	-0.121	-0.132	-0.277	-0.254	-0.112	-0.107
	(0.653)	(0.575)	(0.374)	(0.361)	(0.263)	(0.259)
α_5	1.806	1.338	0.865	0.799	0.634	0.613
	$(0.667)^{***}$	$(0.539)^{***}$	$(0.340)^{***}$	$(0.325)^{***}$	$(0.239)^{***}$	$(0.235)^{***}$
ML	-29.430	-	-73.891	-	-146.017	-
ME	-	35.000	-	77.100	-	159.900
MSE	0.6467	0.6390	0.6739	0.6709	0.1191	0.1192

() standard error is in parenthesis, *** is significant level at 0.05

ML is the maximum value of likelihood function. ME is the maximum value of entropy

Table 1 reports the estimated coefficient of logistic regression conventionally discussed in the logistic function. The calculation involve with the impact of product, price, place promotion and full moon party on the revisit intention of tourists to Pha Ngan Island. We also indicated that the performance of two different estimation methods of logistic regression by applying the sample data of N = 100, N = 200 and N = 400. In the case of N = 100, the estimation of logistic regression of two methods show that price and promotion are significant positive influence to revisit intention. While the GME method perform better than the ML method because the GME method has a small value of MSE. In the case of N = 200, the results illustrate that price and promotion are significant factors to revisit intention and the MSE indicates that the GME method is the better method for estimating the logistic regression. Last, in the case of N = 400, the logistic regression by ML method indicates that the price and promotion are significant positive influence to revisit intention as well as the logistic regression with the GME method. While the ML method has lower mean squared error than the GME method. In addition, we used the standard errors to evaluate the accuracy of the estimated coefficient in logistic regression. We found that all standard errors estimated by GME method were less than the ML method.

5   Conclusions

This study attempts to examine an alternative procedure of estimation method, namely, Generalized Maximum Entropy method in the logistic regression model, in the study of revisit intention of tourists in Pha Ngan Island, Thailand. The results show that the sample of N=100 , N= 200 and N=400 give the results in the same manner. The GME method used to estimate coefficients of the logistic regression gives a lower standard error than the ML method. We used the MSE as a measurement to determine the efficiency and accuracy of a prediction methods. The results confirmed the truth of the MSE from the ML method is lower than the GME method, It is suggested that the GME method may be superior to the ML method, which gives the results similar to the work of [12], in their studied, they used the GME method in order to estimates the parameter in the sample selection model, in the case of a small sample size.

The specific policy implications for improving tourist revisit to Pha Ngan Island, there are two aspects, price and promotion are the two main factors influencing revisit intention. The implementation of an advertisement through various media, such as promotional activities, and selling tour package to tourist directly strategies may prove to be useful as well to make tourists wish to return to Pha Ngan Island again.

References

[1] H. Song, S. F. Witt, Tourism demand modelling and forecasting: Modern econometric approaches, Routledge, 2000.

[2] D. R. Cox, The regression analysis of binary sequences, Journal of the Royal Statistical Society, Series B 20 (1958) 215242.

[3] P. O. Do Valle, A.Correia, E. Rebelo, Determinants of tourism return behaviour, Tourism and Hospitality Research 8 (3) (2008) 205-219.

[4] W. Chaisumpunsakul, P. Pholphirul, Does international trade promote international tourism demand? Evidence from Thailand's trading partners, Kasetsart Journal of Social Sciences (2017) 1-8.

[5] H. Song, G. Li, S. F. Witt, B. Fei, Tourism demand modelling and forecasting: how should demand be measured?, Tourism Economics 16 (1) (2010) 63-81.

[6] O. Claveria, Positioning emerging tourism markets using tourism and economic indicators, Journal of Hospitality and Tourism Management 29 (2016) 143-153.

[7] R. Servidio, I. Ruffolo, Exploring the relationship between emotions and memorable tourism experiences through narratives, Tourism Management Perspectives 20 (2016) 151-160.

[8] E. S. Soofi, A generalizable formulation of conditional logit with diagnostics, Journal of the American Statistical Association 87 (1992) 812816.

[9] A. Golan, G. Judge, J. M. Perloff, A maximum entropy approach to recovering information from multinomial response data, Journal of the American Statistical Association 91 (1996) 841853.

[10] A. Golan, G. Judge, D. Miller, Maximum Entropy Econometrics: Robust Estimation with Limited Data, Chichester, UK: Wiley, 1996.

[11] R. C. Mittelhammer, G. G. Judge, D. J. Miller, Econometric Foundations, New York: Cambridge University Press, 2000.

[12] A. Golan, E. Moretti, J. M. Perloff, A Small-Sample Estimator for the Sample-Selection Model, CUDARE Working Papers in University of California, Berkeley, 2001.

[13] A. Golan, Information and Entropy Econometrics–A Review and Synthesis, New York: Now Publishers, 2008.

[14] A. Golan, Information and Entropy Econometrics Editor's View, Journal of Econometrics 107 (12) (2002) 115.